Department of Education
Office of Federal Student Aid

Data Migration Roadmap:
A Best Practice Summary

Version 1.0

Final Draft

April 2007

START HERE
GO FURTHER
FEDERAL STUDENT AID

Table of Contents

List of Figures

List of Tables

Document History

Change Number	Date	Reference	A M D[1]	Title or Brief Description	Author	Change Request Number
1.0	05-03-07	Final Draft	M	Final changes		

[1] A(dd), M(odify), or D(elete)

Executive Summary

This document is the result of best practice research regarding data migration. It explains what data migration is, what steps are involved, and what common problems and risks one might encounter on a data migration project. This document outlines a practical roadmap to assist with the management of data migration projects.

The Enterprise Data Management (EDM) Team commissioned this document as a service for business representatives involved in data migration projects. This resource is intended to give a high-level overview for those individuals not familiar with data migration and to serve as a reference for those individuals familiar with the topic. This document is written in business language and while there are technical components, it is designed for Project Managers, Subject Matter Experts, and non-technical staff.

Data migration is the transfer of data from one location, storage medium, or hardware/software system to another. Migration efforts are often prompted by the need for upgrades in technical infrastructure or changes in agency business requirements.

A review of best practices found two principles inherent in successful data migration efforts:

1. Perform data migration as a project dedicated to the unique objective of establishing a new (target) data store.
2. Perform data migration in four primary phases: ***Data Migration Planning***, ***Data Migration Analysis and Design***, and ***Data Migration Implementation***, and ***Data Migration Closeout*** as shown in Figure 1.

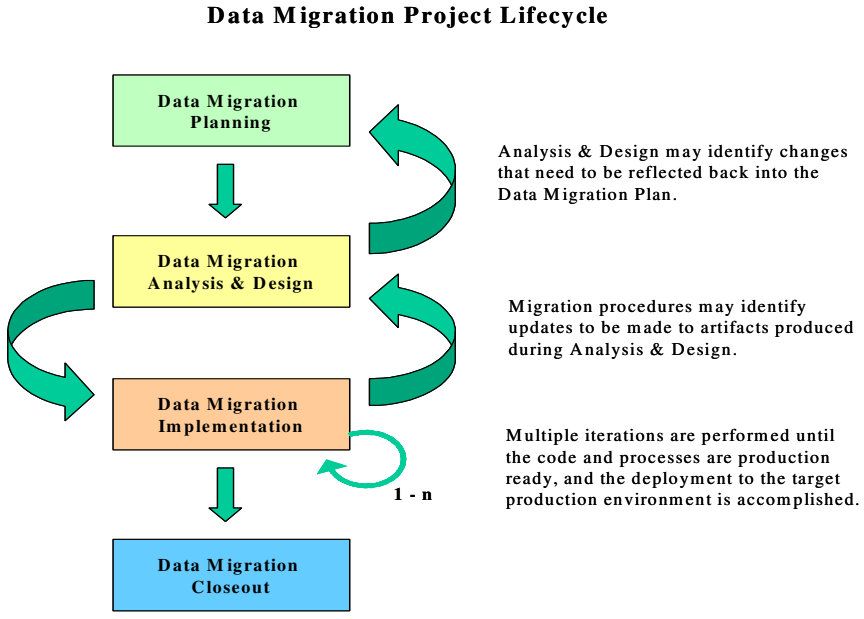


Figure 1: Data migration project lifecycle.

In addition, research found that successful projects were ones that maximized opportunities and mitigated risks. The following critical success factors were identified:

- Perform data migration as an independent project.[2]
- Establish and manage expectations throughout the process.
- Understand current and future data and business requirements.
- Identify individuals with expertise regarding legacy data.[3]
- Collect available documentation regarding legacy system(s).
- Define data migration project roles & responsibilities[4] clearly.
- Perform a comprehensive overview of data content, quality, and structure.[5]
- Coordinate with business owners and stakeholders to determine importance of business data and data quality.

This document is organized according to the four primary phases: ***Data Migration Planning***, ***Data Migration Analysis and Design***, ***Data Migration Implementation***, and ***Data Migration Closeout*** and contains a detailed description of each phase (including tasks and subtasks). In addition, common pitfalls are identified and described. Finally, this document contains a Data Migration Review Checklist, which serves as a tool to help launch and manage data migration projects.

[2] *Microsoft CRM Data Migration Framework,* page 6
[3] *Microsoft CRM Data Migration Framework,* page 8
[4] *Microsoft CRM Data Migration Framework,* page 7
[5] *Strategic Approach to Data Migration,* page 3

1.0 Introduction

1.1 Background and Purpose

Federal Student Aid is engaged in a long-term effort to integrate its processes, data and systems. To better support these business objectives and to emphasize data as an enterprise asset, Federal Student Aid has established a formal Enterprise Data Management (EDM) program. The goal of the EDM program is to consistently define data and make standardized data available across the enterprise by providing information services and data technology expertise to business owners, project managers and architects.

This document is the result of best practice research regarding data migration. It explains what data migration is, what steps are involved, and what common problems and risks one might encounter on a data migration project. The EDM Team commissioned this document as a service for business representatives involved in data migration projects. Many data migration projects differ in scope and complexity. More complex or data-intensive migrations require more elaborate migration plans and procedures. More basic data migration efforts may occur successfully without many of the more extensive procedures and planning steps.

Although simple in concept, data migration can be surprisingly difficult and time-consuming for the following reasons:

- Scarcity of IT professionals with migration expertise.
- Poorly documented source systems with low-quality data.
- Complicated target-system requirements for data validation.
- Complicated development/production environments (running old and new systems simultaneously for a period after the migration).

This document outlines a roadmap and provides checklists to assist with mitigating some of these challenges. Comments or suggestions for improvement to this roadmap are encouraged and should be reported back to the Project Manager for Enterprise Data Management.

1.2 Audience

This resource is intended to give a high-level overview for those individuals not familiar with data migration and to serve as a reference for those individuals familiar with the topic. This document is written in business language and while there are technical components, it is designed for Project Managers, Subject Matter Experts, and non-technical staff.

1.3 Benefits of a Data Migration Roadmap

One way to increase the chances of success on a data migration project is to establish and follow a framework, which is based on best practices and tested and improved through lessons learned. Use of this roadmap, in addition to a project lead's experience, is expected to increase the quality of data migration projects. Specifically, the following benefits are expected:

- **Minimal disruption to the business:** The key to minimal disruption is thorough planning and coordination. Planning should include when and how each stage of the migration effort will occur. In addition, coordination with business stakeholders throughout each stage of the migration is important.
- **Efficient resource utilization (people, budget and time):** Proper data migration planning allows for efficient use of resources. In addition, this planning sets up performance metrics, which can be measured and used to make decisions throughout the migration.
- **Quality Assurance and Risk Mitigation:** Establishing and executing quality assurance and risk mitigation throughout the data migration project improves the chances of a successful project. Proactive quality assurance manages the quality of artifacts and outputs produced and early risk mitigation reduces negative impacts to the project.
- **Cost reduction:** Following a best practice roadmap is expected to result in an overall cost reduction. In addition, collaboration between business owners and technical teams should create minimal business disruptions resulting in reduced cost.

1.4 Document Organization

This document consists of two major sections:

- **Section 1:** Addresses the business needs and benefits of a data migration project.
- **Section 2:** Provides an overview of the data migration project and a detailed description of each project phase at the task and subtask level.

Appendices provide background and reference material about specific aspects of the document, such a technical terms and acronyms. In addition, Appendix G contains selected sections of the document and can be printed out as a "short version" of the document.

2.0 Data Migration Roadmap

2.1 Introduction

A data migration project focuses on the movement of data between legacy (source) data system(s) and a target system, including all necessary procedures for transferring and validating the data throughout the entire process (see Figure 2). Before data is moved, often it needs to be modified and/or transformed. This process is called Data Conversion. Planning and performing data conversion requires the development of transformation rules and procedures to implement the necessary changes. For example, if the legacy system stores date information in text format but the target system requires this information to be stored as date format, then a conversion of the legacy data is necessary prior to the data migration.

It is common to use a staging area as an interim data store to facilitate testing and validation of these modifications/transformations. In addition, a staging area can serve as a storage area for integration projects, which pull data from multiple source systems.

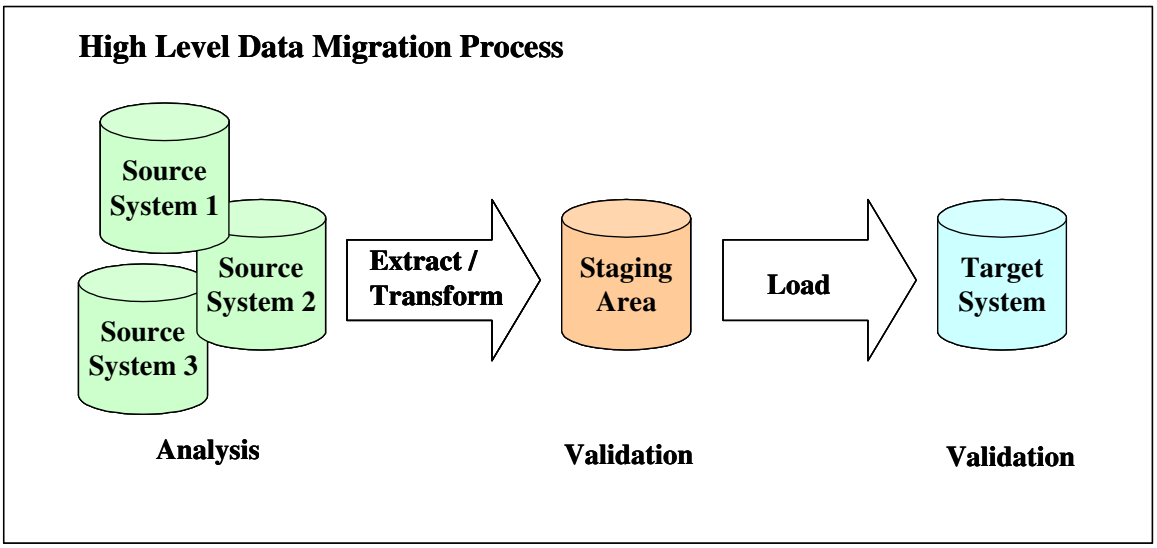


Figure 2: High-level data migration process.

Data migration projects can be initiated for many different reasons and therefore, such projects vary widely in size and scope. The most commonly cited reason[6] for initiating a data migration effort is a technology refresh – the need to upgrade software and/or hardware. Other conditions that often lead to a data migration effort include[7]:

- The need for a more robust system that better meets an organization's technology needs
- A company merger requiring an integration of technology resources

[6] *Simplifying Technology Refresh with Data Migration Software*, page 4
[7] *Strategic Approach to Data Migration*, p 1-3; *Simplifying Technology Refresh with Data Migration Software*, pages 4-16; *The Hidden Cost of Data Migration*, pages 2-6

- The need to resolve data quality shortcomings
- Growth capacity which exceeds original planning
- Loss of a legacy resource (expired lease or license, funding, etc.)
- The development of new data requirements
- Consolidation of data storage/physical storage through hardware upgrade.
- The need to physically relocate data
- The need to increase data availability

A review of best practices produced the following two principles inherent in successful data migration:

1. Perform data migration as a project dedicated to the unique objective of establishing a new (target) data store.

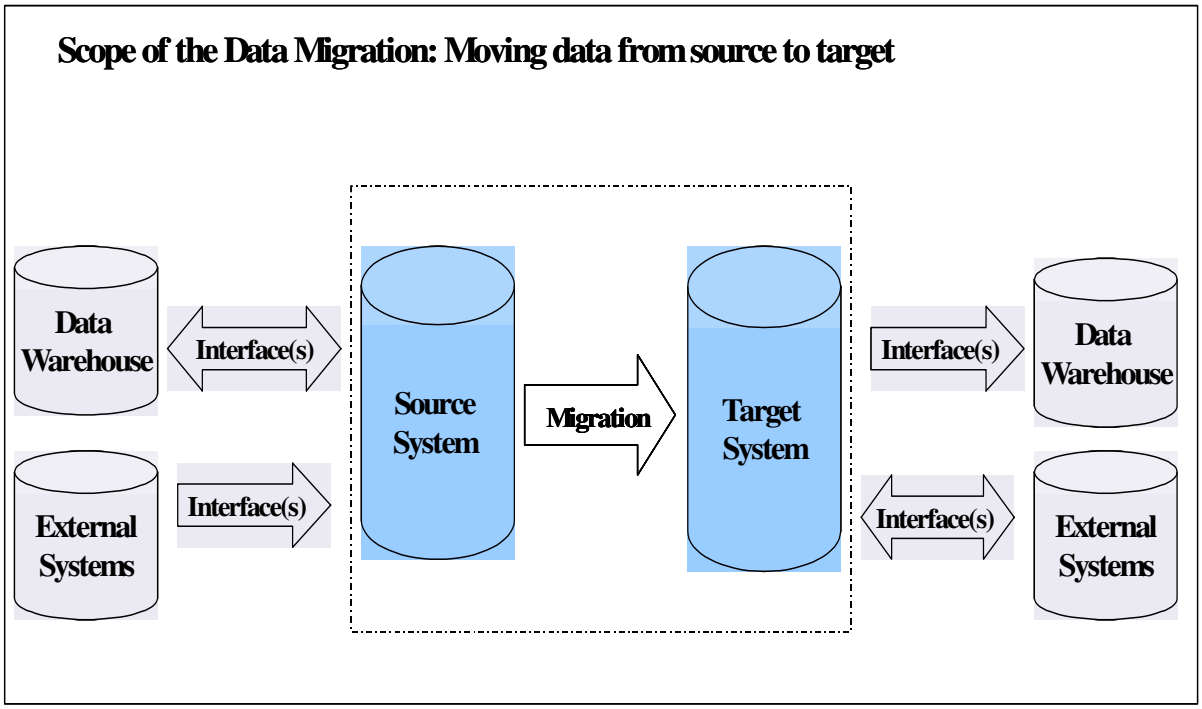


Figure 3: Scope of the Data Migration project

2. Perform data migration in four primary phases as shown in figure 3:
    i. Data Migration Planning
    ii. Data Migration Analysis and Design
    iii. Data Migration Implementation
    iv. Data Migration Closeout

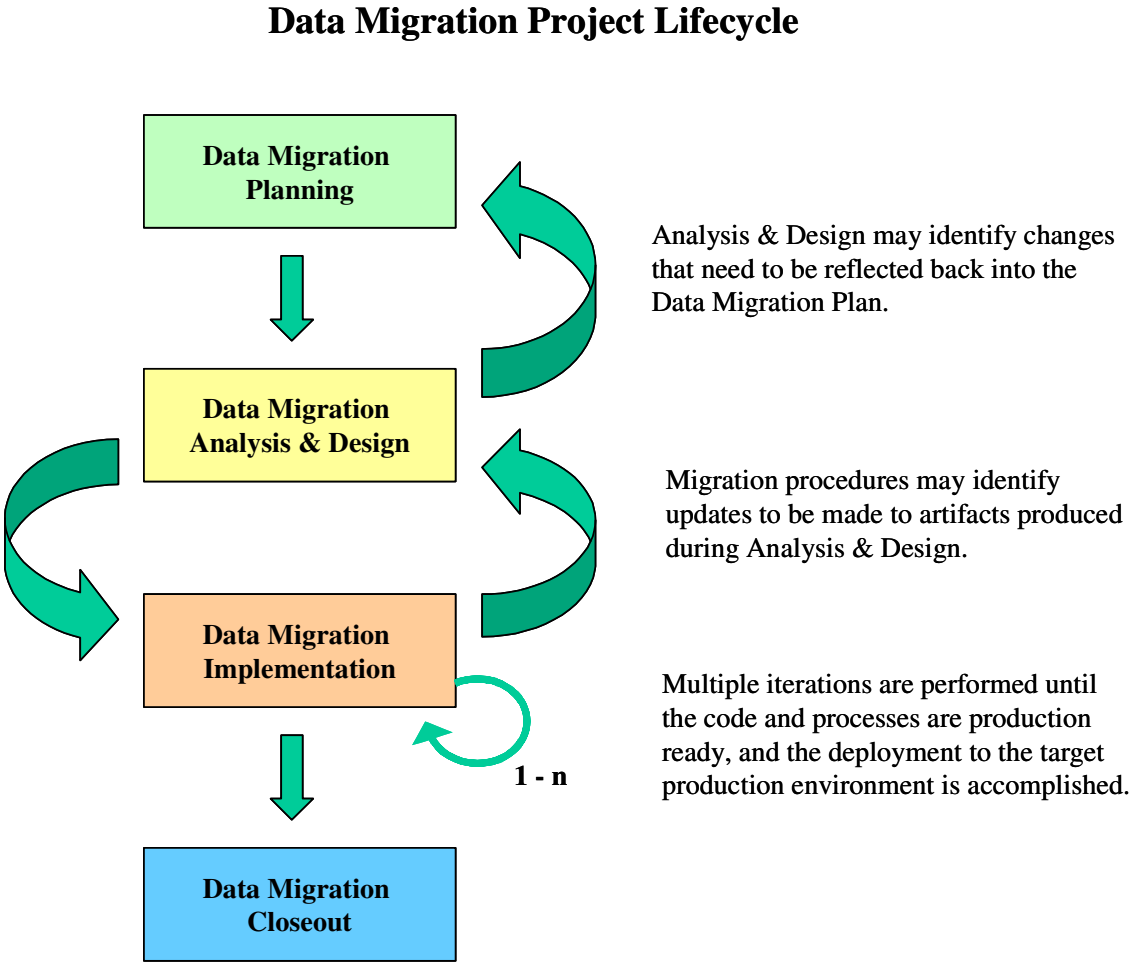


Figure 4: Data Migration Project Lifecycle.

Based on these results of the best practice review, this document recommends leveraging standard project management philosophies such as the Project Management Body of Knowledge (PMBOK). In addition, the document is structured according to the four primary phases.

2.1.1 Data Migration Project Lifecycle

Table 1 lists the high-level processes recommended for each phase of the Data Migration Project Lifecycle. While all data migration projects follow the four phases in the Data Migration Project Lifecycle, the high-level and low-level processes may vary depending on the size, scope and complexity of each migration project. Therefore, the following information should serve as a guideline for developing, evaluating, and implementing data migration efforts. Each high-level and low-level process should be included in a Data Migration Plan. For those processes not deemed appropriate, a justification for exclusion should be documented in the Data Migration Plan.

Data Migration Project Lifecycle			
Data Migration Planning Phase	**Data Migration Analysis & Design Phase**	**Data Migration Implementation Phase**	**Data Migration Closeout Phase**
Plan Data Migration Project	Analyze Assessment Results	Develop Procedures	Document Data Migration Results
Determine Data Migration Requirements	Define Security Controls	Stage Data	Document Lessons Learned
Assess Current Environment	Design Data Environment	Cleanse Data	Perform Knowledge Transfer
Develop Data Migration Plan	Design Migration Procedures	Convert Transform Data (as needed)	Communicate Data Migration Results
Define and Assign Team Roles and Responsibilities	Validate Data Quality	Migrate Data (trial/deployment)	
		Validate Migration Results (iterative)	
		Validate Post-migration Results	

Table 1: Data Migration Project Lifecycle with high-level tasks identified.

During the lifecycle of a data migration project, the team moves the data through the activities shown in Figure 5.

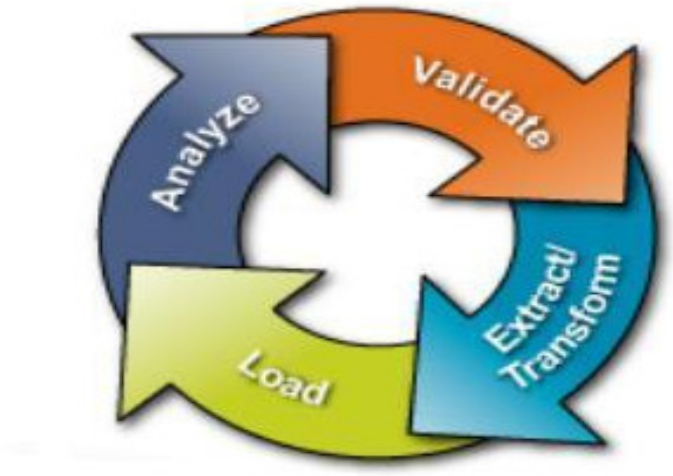


Figure 5: Data Management Activities in Data Migration.[8]

The team will repeat these data management activities as needed to ensure a successful data load to the new target data store.

[8] IPM Data Management Plan (Perot Systems)

2.2 Data Migration Planning Phase

2.2.1 Planning Overview

The Data Migration Planning Phase describes the individual tasks to:

- Plan Data Migration Project
- Determine Data Migration Requirements
- Assess Current Environment
- Develop Data Migration Plan
- Define and Assign Team Roles and Responsibilities

To ensure that both the data migration project and the larger development project are successful, it is good practice to execute data migration as an independent project. Thorough planning is the foundation for consistent success in any process, and data migration is no exception. Also, a successful data migration effort requires the mitigation of issues and risks to the business/ organization.

The *Data Migration Plan* details the information that should be included for each step in the plan. Other results, such as risks and/or critical success factors, may simply be documented in the plan. All steps within subsequent phases of the migration are included in the plan. Also included are the way in which the steps should be performed with respect to rules, parameters, procedures, and so forth.

In addition, the development program describes the general project deliverables, to which all projects must adhere, such as the Quality Plan, Change Management Plan, and Communications Plan. Further, the data migration project might be asked to provide information specific to the data migration effort for the following deliverables:

- Data Quality Plan
- Data Migration Communications Plan
- Data Migration Risk Mitigation Plan

The data migration roadmap covers these particular tasks/deliverables only if there is indeed a data migration specific component that needs to be addressed.

2.2.2 Data Migration Planning Tasks & Subtasks

The following table presents the five major tasks of the Data Migration Planning Phase. Each task is broken down further into subtasks. Subtasks may occur in parallel, or in the sequence shown in this chart. The Data Migration Project Manager, in collaboration with the Data Migration Team (DMT), will determine the order in which these tasks will be performed.

Table 2: Data migration planning.

Data Migration Planning Phase				
Plan Data Migration Project	**Determine Data Migration Requirements**	**Assess Current Environment**	**Develop Migration Plan**	**Define and Assign Roles & Responsibilities**
Establish Scope	Determine Business Requirements & Expectations	Identify and Collect Existing Data-Related Artifacts	Determine Data Migration Method	Define Migration Roles and Responsibilities
Identify Risks/Constraints/ Dependencies /Assumptions	Determine Technology and IT Infrastructure Requirements	Blueprint Current State of the Data Architecture	Determine Conversion Plan	
Develop Data Migration Risk Mitigation Plan	Determine Data Security and Privacy Requirements	Determine Data Migration Technology	Determine Data Integration Plan	
Develop Data Migration Communications Plan			Plan Parallel Operation	
Identify Critical Success Factors			Develop Migration Data Quality Plan	
			Develop Data Archival Strategy	
			Develop Data Migration Test Plan	
Data Migration Planning Artifacts				
Data Migration Planning Checklist				

2.2.3 Plan Data Migration Project

To ensure that both the data migration project and the larger development project are successful, it is good practice to develop a dedicated *Data Migration Project Management Plan* as a subset of the overall *Project Plan.*

Much of the detailed information required by the migration plan is collected during the project, and therefore cannot be included in the initial project plan. This presents two specific issues:

- A combined *Project/Data Migration Project Management Plan* would need to include additional contingencies based on forecasts of the results of the initial planning steps.
- Alternatively, the combined plan would need to be revised to include strategic and tactical information once migration planning is complete. This would be the same as creating a separate Data Migration Plan at the appropriate time.

For this reason, best practices recommend that a data migration plan that is distinct and separate from the project plan be prepared specifically for the data migration effort. The creation of the *Data Migration Plan* allows the capture of results from all previous steps within the actual plan.

Once a plan is drafted, the checklist provided in Appendix C can be used to ensure that all relevant points are addressed. When a step is omitted, a simple justification should be given. The sequence of the steps is a suggestion that needs to be reviewed and adapted for each individual migration. In some cases, steps can be performed in parallel; in other cases, steps must be performed in a particular order. Because all planning steps occur prior to publication of the plan, the plan should recount the results of the planning steps. Some results, such as test plans and data migration architecture, may be referenced as separate artifacts. .

2.2.3.1 Establish Scope

Typically, the scope of a data migration project includes:

- The planning and execution of the transfer of data between a legacy (or source) data storage and a new (or target) storage.
- The design of the supporting structures and functions.
- The procedures for validating the results along the way.

However, a migration project may also include identification of the data source(s) relevant to the migration effort. This information provides valuable insight about the level of effort, the timeline, and the resources needed to accomplish the task. Such information also helps identify dependencies and potential risks.

Data migration projects at Federal Student Aid are often part of a larger development effort. Therefore, the scope of data migration concentrates on moving the data from multiple source systems to a single target system. Modifications related to interfaces and reporting are out of scope for data migration efforts, but would be part of the scope of the larger program of which the data migration effort is a part.

Clear definition of scope at the outset of data migration is important to prevent "mission creep," which might reduce the project's chances of success.

2.2.3.2 Identify Risks/Constraints/Dependencies/Assumptions

Each data migration project entails potential obstacles: specifically, risks, constraints, challenges, dependencies, and assumptions. While planning the stages of the data migration effort, identify as many of the obstacles as possible. Thorough planning forces the data migration team to consider and anticipate a broad range of events and results that may occur or that may be prevented during the data migration. The better these obstacles are managed, the better the chances become for a successful data migration.

Every data migration presents a unique set of issues and risks that must be monitored. Table 3 on the next page shows a preliminary list of typical assumptions/dependencies, constraints, and risks to consider:

Table 3: Assumptions, constraints, and risks.

<table>
<tr><th>Assumptions/Dependencies</th><th>Constraints</th><th>Risks</th></tr>
<tr><td>Sufficient resources are available for all aspects of data migration.</td><td>Time/Schedule dictates what must be completed and when.</td><td>Unexpected delay and/or downtime might occur.</td></tr>
<tr><td>Sufficient expertise is available for all aspects of data migration.</td><td>Funding might limit access to resources that can be devoted to the effort.</td><td rowspan="2">The team might encounter complex:
i. Processes
ii. Environments
iii. Configuration issues related to data volumes</td></tr>
<tr><td>All environments (legacy, staging, target) are fully documented, available, and accessible as planned during necessary steps of migration.</td><td>Personnel and equipment might be limited or unavailable.</td></tr>
<tr><td rowspan="2">The team has access to:
○ Subject matter experts for current source system and data
○ Documentation for data models, physical implementation (database), and business rules, and interfaces
○ Future business requirements</td><td>Data requirements and definitions might require clarification by subject matter experts.</td><td>Misunderstanding or misinterpretation of requirements might result in a flawed design.</td></tr>
<tr><td>Expertise in legacy-storage environment might be limited due to lack of or outdated documentation.</td><td>The team might encounter incompatible software, hardware, processes owing to:
i. Multiple Operating Systems (OS) or vendors
ii. Format incompatibilities (Database Management System (DBMS) to DBMS, DBMS to Operating System, etc.)</td></tr>
<tr><td rowspan="2">All selected tools and software packages necessary for data migration will be available and implemented for the necessary migration steps as outlined in the overall project schedule. In addition, the necessary licenses will be available to the project team.</td><td>Availability of current and target physical storage (lease conditions, support, physical condition of hardware/software, etc.) might be limited.</td><td>Expensive overtime might be required to do certain steps during non-business hours to reduce impact on production.</td></tr>
<tr><td></td><td>Unplanned events or conditions might occur (e.g. configuration of target system is delayed due to illness of support staff.)</td></tr>
<tr><td colspan="2" rowspan="4"></td><td>There might be problems with physical relocation of hardware or data.</td></tr>
<tr><td>Legacy data architecture artifacts might be unavailable or incomplete.</td></tr>
<tr><td>The new data store might be physically incompatible with the legacy location (such as hard drive connection to server or device drivers).</td></tr>
<tr><td>The schedule might slip owing to slow or delayed acquisition process of migration software or equipment (e.g., server) and delayed availability.</td></tr>
</table>

2.2.3.3 Develop Data Migration Risk Mitigation Plan

During this data migration planning phase, the Migration Management Team (MMT) must develop a *contingency strategy* (alternate method of accomplishing the same objective) or *risk mitigation plan* (a means of reducing the impact of undesired results) for each anticipated risk

that could jeopardize the successful completion of the migration effort. It is important to continue proactive issue management and proactive risk management throughout the lifecycle of the project. Each data migration project deals with different issues and risks, and therefore each requires a Risk Mitigation Plan specific to the scope of the project. Similar issues and risks may have been identified and/or have occurred during other data migration projects of which the team has knowledge, and thus the mitigation and resolution approaches from those projects should be used as guidance. However, a previously-successful solution may not apply to the current project, and therefore, such a solution should be carefully evaluated. This is the main reason why each individual project requires its own specific Risk Mitigation Plan.

Note that this Risk Mitigation Plan needs to align with the Risk Mitigation Plan of the overall development project.

A successful data migration effort requires the mitigation of issues and risks to the business/ organization. Identifying these challenges (such as dependencies on other teams within Federal Student Aid, minimal migration expertise within the organization, or insufficient understanding of data and source systems) and opportunities (such as the identification of the most appropriate data migration method) early in the project allows for proper management and less later disruption. In order to decrease risks, project leads should consider the following actions:

- Migrate only the data required to sustain the future application
- Identify and employ the most appropriate migration method to move the required data into the new solution
- Ensure that all data is migrated accurately and completely
- Ensure that the integrity of the migrated data is maintained
- Minimize disruption to the business during transition
- Prepare a detailed inventory of what data and systems architecture exist, and identify any data issues relevant to the conversion during the early phases of the project
- Identify any resource dependencies, such as access to and availability of environments (source, staging, target system), tools, software licenses, or personnel. Constraints and restrictions of the target system may require the development of complicated data validation procedures to ensure the integrity and quality of the data loaded.
- Identify staff/resources with knowledge of and experience with the source data. This will reduce the risk of undocumented data issues and will allow identification of potential pitfalls and other issues.
- Identify potential challenges and opportunities early on in the project. This allows for proactive management of these challenges and opportunities.

Risk mitigation is an important task that covers the entire project lifecycle. It is also important to consider that issues and risks detected and addressed in the planning and design phase of a project are less costly than those discovered during the implementation phase.

Best practices indicate that the best way to mitigate risk during a data migration is three-fold:

- Employ commercial data migration software (data profiling, ETL, metadata management, etc.)
- Educate management and technical staff about the features and availability of data migration software
- Reduce costly downtime by selecting an online data migration method[9] (See section 2.4.12)

Even with the most thoroughly tested tools and procedures, whether commercial or custom-built, the conditions necessary for a successful data migration may not occur because they are outside of the control of the data migration project. For instance, if the data migration environment is not configured as per the specifications provided by the Technical Migration Team (TMT), if the team is unable to execute the extract and transformation procedures in this environment, or if there is an insufficient number of software licenses available, the execution of a particular task can only be performed sequentially instead of in parallel. This will prolong the originally-planned duration of this task. These conditions, and the likelihood of their occurrence, should be documented as thoroughly as possible and evaluated through a *Risk Mitigation Matrix*.

This *Risk Mitigation Matrix* should be the standard mechanism for reporting risks and their corresponding contingencies or mitigation solutions. Each identified risk should be given a probability rating and impact level, and a brief statement should be made of a potential mitigation solution. Table 4 shows probability ratings.

Table 4: Data migration risk probability and impact levels.

Probability Rating	Impact Level
High	Likelihood > 70%
Medium	40% < Likelihood < 70%
Low	5% < Likelihood < 40%

Impact levels are:

- Catastrophic: failure of mission-essential services
- Critical: significantly degraded project performance
- Marginal: significantly degraded support function or secondary mission
- Negligible: inconvenience

Table 5 shows selected risks and risk-mitigation strategies, including their probability ratings and impact levels.

[9] *The Hidden Costs of Data Migration*, page 6

Table 5: Data migration risk mitigation matrix.

Risk Description	Probability Rating	Impact Level	Mitigation Strategy
A Physical Data Model (PDM) does not exist for each legacy system. As a result, the precise structures are unknown to the implementation team, and the schedule is delayed while the missing PDM(s) is/are reverse engineered.	High	Critical	Work closely with Federal Student Aid and legacy contractors to gather PDM documentation early. Upon identifying any missing PDMs, the required database schema shall be reverse-engineered to provide the necessary PDMs.
Data quality issues are not identified until late in the project, thus causing delays and cost overruns.	Medium	Critical	Quality review sessions will be conducted throughout each release so that data quality issues may be identified early and addressed accordingly.
Necessary database personnel are not available during migration.	Low	Critical	In case access to Federal Student Aid resources is very limited, the contractor should consider hiring a short-term consultant to develop the databases to support the target data.

Issues arising during the life cycle of the data migration project need to be reported, documented, and resolved as soon as they arise. At a minimum, the procedure for handling a concern/risk/issue should include the following steps:

- Identify the concern/risk/issue (documentation).
- Communicate the concern/risk/issue to the team.
- Assign responsibility for analysis and resolution.
- Analyze and determine a resolution or mitigation strategy (or explain inability to resolve).
- Seek approval of the mitigation and/or solution.
- Implement the approved solution, if possible.

Review the *Data Migration Risk Management Plan* should be reviewed on a regular basis to ensure appropriate monitoring of risks.

2.2.3.4 Develop Data Migration Communications Plan

The *Data Migration Communications Plan* identifies all data management aspects (what, who, when, where, how, about) of the data migration project to stakeholders, Data Migration Team members, and (if needed) external personnel. This plan outlines the recipient, title of

communication, content, format, and schedule of each document prepared and shared as a result of the data migration project. Only communications relevant to the migration effort are discussed. The *Data Migration Communications Plan* should cover at least the following:

- Status reports (weekly, monthly)
- Deliverables and their distribution list including approval authority
- Escalation procedures
- Data profiling findings
- Data cleansing findings
- Information about migration specific metrics, such as:
    - Data volume in source systems and what this means to the project with respect to throughput capability, time to migrate, etc.
    - Performance for data extract/load procedures
    - Sizing of staging area and target system
- Trial migration execution(s) results, including benchmarks
- Important decisions made or reviews passed throughout the lifecycle of the project, including:
    - Stage gate reviews
    - Approval for final data migration (deployment to production)
- Details about the final migration to target system (deployment) including metrics such as:

Data volume migrated (total number of records) and statistics on:

- Accuracy (correct and flawed records in total number and percentage)
- Completeness (completed and failed records in total number and percentage)
- Total time to migrated all records
- Downtime of production system (start to finish)
- Throughput

- Post-migration activities, including:
    - Validation activities occurred (e.g. test cases performed and results)
    - Steps to resolve remaining migration issues (fix flawed records, research and verify any lost data, etc.)

These communications, and their content, delivery schedule, and format of delivery (electronic or paper), will be outlined in the *Data Migration Communications Plan*.

2.2.3.5 Critical Success Factors for Data Migration Planning

The MMT, in collaboration with the stakeholders, must identify critical success factors for the data migration project. These success factors are the elements considered crucial in ensuring that the data migration effort attains its objective of thoroughly, cleanly, and efficiently transferring the business data from the legacy data environment to the target data environment.

Best Practices identify the following critical success factors while planning a data migration effort:

- Understand source and target data requirements and structures.
- Define project roles and responsibilities.
- Provide a comprehensive overview and accurate insight into data content, quality, and structure.[10]
- Document and discuss any anticipated issues or risks with stakeholders and/or business owners.
- Perform migration as an independent project.[11]
- Establish and manage expectations throughout the process.
- Understand current and future data and business requirements.
- Identify individuals with expertise regarding legacy data.[12]
- Collect available documentation regarding legacy system(s).
- Clearly define data migration project roles & responsibilities[13].
- Prepare a comprehensive overview of data content, data quality, and data structure.[14]
- Determine the importance of business data and data quality with business owners and stakeholders.

Each data migration effort, however, has its own specific factors. Their determination and documentation are the responsibility of the MMT.

2.2.4 Determine Data Migration Requirements

Requirements for data migration projects are as varied as they are critical. Outlining the business requirements for the data helps determine what data to migrate.[15] These requirements may take

[10] *A Strategic Approach to Data Migration,* page 1
[11] *Microsoft CRM Data Migration Framework,* page 6
[12] *Microsoft CRM Data Migration Framework,* page 8
[13] *Microsoft CRM Data Migration Framework,* page 7
[14] *Strategic Approach to Data Migration,* page 3
[15] *Microsoft CRM Data Migration Framework,* page 6

the form of any necessary agreements, expectations, and/or objectives of the migration.[16] All information is captured in the *Data Migration Requirements* document.

The legacy environment is generally static during a migration, so the Assess Current Environment step (**Section 2.2.5**) should encompass the operational/technical requirements of the current environment. Any requirements for synchronizing changes in content or structure of the legacy environment during the migration must be defined.[17] The Technical Lead of the MMT must describe in detail any operational/technical requirements for the target and interim (staging area) environments.

2.2.4.1 Determine Business Requirements and Expectations

The MMT consults the business-area stakeholders and subject matter experts (SMEs) regarding any requirements they might impose above and beyond the technical requirements for the data migration.

Together with the business-area stakeholders, the MMT must:

- Establish requirements to be supported by the structural and procedural designs, including:
  - Iterative or phased approach
  - Standard format for all artifacts (exposing metadata for validation (mappings and data element dictionary))
  - Migration/replication design requirements
  - Volume of data
  - Physical relocation of data storage during or after migration
  - Required application performance before, during, and after migration
- Establish desired timelines for the individual elements of the data migration, including:
  - Schedule
  - Availability of current and target physical storage (lease-related, support, condition)
  - Hardware availability
  - Allowable downtime during migration
- Identify all business processes that will be used and/or affected by the proposed data migration
- Consider future requirements (build for growth and scalability)
- Determine stakeholder requirements

[16] *How to Plan for Data Migration (ComputerWorld, May 21, 2004)*
[17] *Taking the Pain Out of Data Migration,* page 1

- Identify relevant business rules and processes
- Identify existing data migration related artifacts for the current environment and approve creation of necessary artifacts where shortcomings are recognized. A technical team will be required if technical artifacts must be created.
- Review Federal Student Aid data security policies and determine applicable security requirements for the data migration

The MMT must also coordinate with any anticipated user community, such as users of the affected software and/or hardware, in addition to other users of the migration technology. The purpose of such coordination is to identify expectations users may have about the effect of the data migration before, during, and after execution.

2.2.4.2 Determine Technology and IT Infrastructure Requirements

As discussed in the introduction, this document presents a tool-independent methodology. However, nearly every aspect of a data migration effort can be automated to some degree, so technology cannot be ignored. It is critical for all of the pieces of a data migration to work together in order successfully to move, cleanse, and/or convert the legacy data to a new environment. Therefore, the technology used at each point in the process should be described in the *Data Migration Plan*.

Together with the business-area stakeholders, the MMT must establish technology requirements. This means defining the conditions and objectives to be satisfied by whatever technology is eventually chosen. A clear understanding of the migration effort will help determine the best technology to use for the migration.[18] Areas to consider are:

- Data Migration technology infrastructure that is in place, or standard products endorsed, such as:
  - Data file format, management software and database management systems (such as XML, Sequential File, MS SQL Server 6.0, MS SQL Server 2005, and Oracle 10g) of the legacy, staging, and target data environment
  - Operating system (such as UNIX, Linux, MS Windows 2003) of the legacy, staging, and target data environment
  - Interfaces
  - Network and communication environment
  - Extract, Transfer, Load (ETL) software
  - Modeling tools (System Architect (formerly Popkin), Metis, ProVision, ERwin, Embarcadero (ER/Studio, etc.))
  - Data profiling software
  - Metadata management software (other than architecture-design software)

[18] *2006 Best Practices for Data Migration,* page 7

- Hardware configuration
- Distribution of data storage
- Homogeneous versus heterogeneous storage requirements
- Software configuration
- Additional requirements for a multi-vendor storage or software environment
- Performance requirements of procedures/software

For example, ETL software might execute slightly differently against an Oracle database under a UNIX operating system than it does against a Microsoft SQL Server database under a Windows operating system. Such software might not be compatible with the architecture of the staging area. For these and many other reasons, a description of the technology involved in any migration should be prepared.

Best Practices often include an activity that is technically outside the scope of the actual data migration. Considering the requirements of the target data store with regard to longevity and future activity can significantly affect decisions made about and during the migration. It is generally reasonable for the MMT to consider future requirements of the target data store such as:

- Durability of target data environment
- Migratability of target data environment in the future
- Re-usability of target data environment
- Scalability of target data environment
- Anticipated future data capacity
- Additional future applications access the target data environment

2.2.4.3 Determine Data Security and Privacy Requirements

The MMT must review and follow the processes and roadmaps outlined in the *Handbook for Information Assurance Security Policy Information Assurance Program March 31, 2006* for protection of all data at each source, as well as during the migration of the data between sources.

2.2.5 Assess Current Environment

The assessment of the current environment requires the compilation of all identified artifacts to create a blueprint of the current (legacy) data architecture. In the event that the inventory of data architecture artifacts is incomplete and/or insufficient to address the entire legacy environment, a technical team must complete the architecture.

2.2.5.1 Identify and Collect Existing Data Related Artifacts

The outcome of this step is the foundation for the data and systems architecture of the data migration project. Documenting and analyzing the current environment from a functional and technical perspective is important for full understanding the data and the related business rules and processes. Creating such documentation and performing such analysis can be time consuming, depending on the ease of access and the availability of any required information. This inventory of facts drives the development of the data migration procedures. Following is a list of sample artifacts that need to identified and collected:

- Information/data architecture and system architecture documentation of source systems, such as:
    - Logical and physical data models (entity-relationship diagrams and/or repository information of the data structures)
    - Database definition language (DDL) for existing relational databases
    - Data dictionaries documenting each data element (labels and definitions as well as properties)
    - Relevant business rules and processes in the current environment and for the future target system
    - Data mapping from source system to data warehouse
    - Names of systems interfacing with the source systems indicating whether the application sends data to or extracts data from the source system (for context only). This information will help gauge the time constraints for potential downtime of the source system. If there are many interfacing systems, the coordination task will be more complex than with fewer interfaces
    - Data profiling analysis for source system (if available)
- Information about known technical constraints of the source system that affected implementation decisions (e.g. limitation on throughput or performance of the server or software version that did not support originally-planned technical solution)
- Information about any known issues or concerns regarding the quality of the available documentation, such as whether the documentation was outdated (e.g. documentation was prepared when the original project started 5 years ago). This information can only be collected through interviews
- Information about any known and identified gaps/missing information that should be resolved after the data migration, which may or may not be documented as a business or technical requirement. This information can only be collected through interviews

2.2.5.2 Blueprint Current State of the Data Architecture

The information collected during the previous task allows for the preparation of a blueprint of the current state of the data architecture for each source system. This blueprint focuses on the logical data model and the data dictionary, if available. If documentation exists only for the physical implementation, this information must be reverse engineered, resulting in a logical data model. The logical data model presents the relationships and the business context of the information used in the application. The physical data mode, by contrast, demonstrates the implementation of the data from a technical perspective, meeting performance and data access path requirements. This data is essential input to the success of the data migration project.

2.2.5.3 Determine Data Migration Technology

Available resources and the migration method(s) selected for the effort will determine most of the technology used for the data migration. As a first step, the technology established by Federal Student Aid, such as the architectural design tools and metadata repositories, should be evaluated to determine whether they meet the requirements. Next, a report stating the shortcomings and the corresponding requirements should be prepared and presented to Federal Student Aid for review and decision on how to proceed, if the current technology requirements cannot support all requirements.

The MMT will work closely with the DMT to identify the IT infrastructure required to implement the data migration efforts, as well as and any infrastructure affected by the execution of the proposed data migration.

Existing technology, such as the source-data storage devices and software, are already in place, requiring no decision, but rather must be captured as part of the Legacy (or baseline[19]) Data Architecture. However, the tools in place to *access* the legacy data during the migration (such as ETL software or custom programming[20]) should be evaluated to ensure migration requirements are met. All technical aspects of the staging area (if used), target environment, actual movement, and validation of the data must be defined and documented by the TMT.

Aspects to consider when evaluating and/or selecting the most appropriate migration software:

- Criteria to validate existing software:
    - Does the migration software facilitate restoring the pre-migration state in the event of errors or failure?
    - Does the migration software support differences in volume size between source and target data?
    - Is the migration software compatible with both the source and the target storage hardware?
    - Does the migration software support online data migration, if this method was chosen?
    - What are the rollback capabilities of the product?
    - Does the migration software support non-disruptive data migration (i.e., legacy application continues to operate during the migration)?

- Additional criteria to be considered, if software needs to be purchased:
    - Does the intended operating system support this migration software?
    - Does the intended hardware support this migration software?
    - How quickly is data transfer performed?
    - Can the operational environment accommodate the processing requirements of the software?

[19] *A Practical Guide to Federal Enterprise Architecture, February 2001 (CIO Council),* page 5
[20] *The Complete Data Migration Methodology*, page 6

- Is it cheaper to custom-build the migration tools than to buy?
- What is the acquisition timeline, and will it cause a delay in the project?
- How quickly can the new software be implemented and staff trained?
- What is the ramp-up time from a resource perspective?

As part of this analysis, the DMT will determine whether the IT infrastructure in place will support the planned data migration effort and, if not, what solution to recommend resolving any shortcomings.

2.2.6 Develop Data Migration Plan

As with any IT project, communication is critical to success[21]. The MMT must compile the results of all planning steps and draft the *Data Migration Plan.* The plan shall be submitted to the EDM Team for evaluation. Once the plan has been evaluated for consistency and compliance with enterprise expectations, the plan should be presented to all affected business area owners and stakeholders for feedback and approval.

After all revisions have been finalized and approved, the plan should be provided to the TMT for implementation.

2.2.6.1 Determine Data Migration Method

Technology is constantly changing[22]. Whether automated or manual, six basic methods of migrating data are published. These are generally divided into two broad categories based on whether the procedures may be performed while the application remains operational (online) or the application must be taken out of service (offline) during the actual migration. The basic methods are[23]:

- Offline: back up & restore; restore from backup tapes; ftp transfer, and
- Online: array-based replication; volume management or replication; and host-based mirroring

In many cases, a hybrid of these methods is required to satisfy the requirements of a major migration effort. The Data Migration Project Manager, in close collaboration with the Federal Student Aid EDM Team, must determine the method and tools to be used to perform the activities of the data migration. The method can differ based on the legacy systems involved in the data migration. Also, not every migration method is suitable for every source and/or target data environment. The method and tools chosen, and the factors contributing to the determination, should be included in the Data Migration Plan. Such factors may include (but are not limited to):

[21] *How to Plan for Data Migration (ComputerWorld, May 21, 2004)*
[22] *Data Migration: Plan to Succeed (DM Review, January 2007)*
[23] *Simplifying Technology Refresh with Data Migration Software,* page 6

- Distribution (location) of data stores
- Funding constraints
- Available expertise in current and target storage environment (e.g., whether planning is limited to specific options simply because of available expertise)
- Performance (qualitative and quantitative) of procedures/tools
- Source data protection/recovery
- Homogeneous versus heterogeneous storage requirements
- Multi-vendor environment
- Dependencies on external business partners
- Allowable downtime
- Time (schedule) constraints
- Volume of data
- Personnel constraints (availability)
- Complexity of storage and processing environment
- Physical re-location
- Data storage format incompatibilities (DBMS/DBMS, DBMS/OS)
- Configuration issues related to data volume

2.2.6.2 Determine Data Conversion Plan

Federal Student Aid requires the MMT to develop a plan for satisfying such requirements.

Migration requirements may require a change to the legacy data during the migration process. The changes may be to form, value, or volume. A strategic approach to data migration that analyzes legacy data at the source will mitigate this risk[24] by allowing analysis both at the source and at each step of the migration process. The MMT, and specifically the data stewards, must define the form and business function of the target data. Transformation of the data values, constraints, and/or format occurs during the migration process, through thoroughly tested rules and procedures. The following sample requirements can trigger data conversions:

- Conformance to target data requirements
- Translation of source data values to target data values,
- Enhanced data types and/or formats in newer technology

2.2.6.3 Determine Data Integration Plan

Data migration may require drawing data from more than one legacy data source. The *Integration Plan* describes how conflicts and duplication in source data and data structures will be resolved. The plan also determines how to move the data from the source system(s) to the target system. There are two options:

- Load the data sources sequentially in to the Staging Area until all source data has been loaded. Then, perform the integration of all source data in the Staging Area. Finally, move the integrated data to the target data store.

[24] *A Strategic Approach to Data Migration,* page 1

- In some cases the volume of data or time restrictions may not support the above-described option, and may result in sequential individual data migrations (one for each source system). The staging area could serve as an integration environment to simulate loading the new data set into an environment already populated with operational data.

Staging areas are an optional interim data source, which can serve the purpose of mirroring the ultimate target system for many data migration efforts. Best practices demonstrate the benefits of establishing a staging area. Loading data into the staging area supports the simulation of the data load into the target system. It allows validation, cleansing, and/or conversion of the integrated data prior to movement into the target location. These trial migrations can be repeated multiple times until the data migration procedures are perfected without affecting the configuration and readiness of the final target system.

The need to integrate multiple legacy-data sources mandates such a staging area.

2.2.6.4 Plan Parallel Operation

Migrating financial systems often require the old and the new system to run in parallel for a pre-defined period of time to ensure the reliability and accuracy of the newly implemented target system. Federal Student Aid follows this principle when planning whether the legacy systems should continue operation for a set time after a successful migration. The legacy system may even serve as a long-term data source for the target system (which is often the case when migrating data from an operational, or transactional, system). However, some legacy systems may be scheduled for complete shutdown upon successful migration of the data to a target system. Others may already be out of operation, which may be the leading factor facilitating the migration.

While different purposes are served by shutting down or continuing operation of legacy systems, the two scenarios have one issue in common: both require that the data contained in the source *and* target systems remain synchronized to some degree as long as both systems are in operation. In the case of a transactional system feeding a data warehouse, the source data is often derived and/or aggregated over a particular time period when being moved into the warehouse. These rules and algorithms shall be developed as part of the procedures for populating the target data store.

A third scenario involves maintaining the legacy data store for a period of time while the operation of the new system is validated. While this is generally considered a post-migration task, the full operation of the new system may reveal errors in the migrated data, requiring revision to some part of the data migration (data quality remediation, data migration procedures, etc.). Roadmaps for operating the two systems in parallel, monitoring and comparing the performance of each system, and resolving issues as they arise should be established and included in the *Data Migration Plan*.

2.2.6.5 Develop Migration Data Quality Plan

The *Data Quality Plan* concentrates on the quality of the legacy data. It asks for multiple efforts that can be performed in parallel. All outcomes will determine the overall data quality of the source system(s) to be migrated. In addition, the *Data Quality Metrics* and *Data Loss Tolerance* information shall be used as benchmarks to determine the fitness of the data for deployment.

Define Data Quality Metrics: A *Proof of Concept*, which simulates a full data migration by operating on a sampling of data supporting a single event, such as a single transaction or single concept[25] may be performed if a commercial data migration (such as ETL software) or data profiling tool is used. The *Proof of Concept* provides a field-level and/or record-level view of the legacy data and helps identify anomalies. The document will validate the compatibility of the technology selected to perform the data migration, and will provide data quality metrics based on the sample data that may be used to project the level of effort required to perform full data remediation. If custom software or procedures are planned for the data migration, then data metrics must still be established for measuring the quality and integrity of the data before, during, and after each migration stage.

Define Data Loss Tolerance: If all data stores that participate in a data migration effort (legacy, staging, and target) have the same basic specifications (e.g., a relational database using version X of RDBMS Y on operating system Z, etc.), it is reasonable to expect that all data will transfer without loss. However, on occasion obsolete data structure or formats may not translate 100% into a modern environment. The MMT must consult the business stakeholders to determine the tolerance level for data loss. This will contribute to the remediation (what to do if a level of data loss is unavoidable) and the validation procedures (measuring whether the migration is successful).

Establish Data Quality Remediation Plan: The creation of a "zero-defect" data quality policy is optimal prior to data migration. Such a policy can be put into place by performing error correction, including duplication (at the source (passive) or through an active remediation process), which corrects data errors *during* the migration process. If this cannot be accomplished, then fixing known or discovered errors in the legacy data should be the first post-migration step[26]. Once metrics are established, the MMT must determine at what stage of the migration, and by what means, the identified data quality issues shall be remedied, and lay these decisions out in the *Data Quality Remediation Plan*.

If a passive remediation plan is chosen that affects the content of the source (legacy) data, notification to dependent systems of changes must be included in the *Communications Plan*. When planned properly and included as part of the migration process, data remediation ensures the integrity of the end product (target data).

2.2.6.6 Develop Data Archival Strategy

The plan for managing data once it is no longer necessary for immediate access and use is called the *Data Archival Strategy*. The MMT must interview stakeholders and formulate strategies regarding what data to retain, how long to retain it, and where and how to retain it.

---

[25] *A Strategic Approach to Data Migration,* page 3
[26] *A Strategic Approach to Data Migration,* page 1

A thorough architecture of the legacy system may already include a *Data Archival Strategy*, but it likely only covers the retention and management of data during the operational life of the legacy system.

If the strategy in place for the legacy system is sufficient to address the retention of data once the system is removed from operation, then the full strategy may be adopted as part of the *Data Archival Strategy* within the *Data Migration Plan*. If, however, the strategy does not address system shutdown or does not exist at all, then the MMT must establish a strategy for retaining and retrieving the legacy data after the legacy system is taken out of operation, if necessary. This may include a period of parallel operation with the target system.

Since the migration staging area and target data store are likely being created as part of the migration, the MMT must establish the *Data Archival Strategy* for both of these data stores after consulting with the stakeholders. The combined strategy should guide the creation of the applicable data architectures during the design steps.

2.2.6.7 Develop Data Migration Test Plan

The MMT must establish a plan for testing the migration procedures at each step along the way. All data movement procedures, transformation/conversion procedures, data cleansing procedures, and data validation procedures must be accounted for in the context of the *Migration Data Architecture*. The data migration procedures must be able to successfully satisfy the requirements set forth in the data requirements and this plan before proceeding to the full migration.

2.2.7 Define and Assign Roles & Responsibilities

2.2.7.1 Define Migration Roles and Responsibilities

The proper procedures for establishing and documenting the roles and responsibilities for a data migration project are no different than for any other information technology project. The *specific* roles and responsibilities, however, are often unique to a data migration project. It is recommended that roles and responsibilities be established as early as possible during the planning stages in order to help guide the DMT throughout the project[27].

Table 6 shows the data migration project's specific roles and responsibilities for stakeholders and team members.

[27] *Microsoft CRM Data Migration Framework,* page 7

Table 6: Recommended Roles & Responsibilities.

Data Migration Roles	Responsibilities
Stakeholders (Federal Student Aid Business Owners, Enterprise Architects, etc.)	Provide guidance and input to the overall project Determine acceptance and success criteria for data migration Approve the *Data Migration Plan* and other artifacts prepared by the data migration project team Approve production-readiness of migration procedures and the timeline for deployment of data migration.
Enterprise Data Management (EDM) Migration Team (May include Program Manager, Data Architects, Governance Team, etc.)	Provide guidance to data migration teams in establishing *Data Migration Plans* Evaluate the *Data Migration Plan* prior to approval by the Stakeholders and implementation by the Technical Migration Team Provide the Technical Migration Team with any standard enterprise metadata (standards, policies & procedures, designs, etc.) necessary to develop the *Data Migration Architecture*
Migration Management Team (MMT)	
Project Manager[28]	Plan the project (but not the migration unless serving double duty as Migration Manager) Execute the project plan Monitor scheduling, progress, performance, and issues and risks Close the project
Migration Manager (Could be the same as Project Manager based on scope/budget of the project)	Prepare the Data Migration Plan and plan the data migration in collaboration with Project Manager and data migration project team Manage and monitor the Implementation & Validation of the data migration Collaborate with the Project Manager on risk mitigation Document data migration results Report to the Project Manager

[28] PMP In Depth, page 8

Data Migration Roles	Responsibilities
Data Steward(s)	**Planning:** Compile legacy data architecture, including models, data dictionaries, volume metrics, and other artifacts **Analysis & Design:** Coordinate / validate data designs and inter-environment (source / staging / target) correlations; facilitate development and validation of data profile metrics **Implementation:** Coordinate data cleansing and data quality remediation **Closeout:** Coordinate validation of reports on the results of the data migration
Migration Technical Lead (TML)	Provide technical expertise to the Migration Management Team during Planning Compile and document technical requirements during Planning and Analysis & Design Lead the Technical Migration Team during Analysis & Design, Implementation, and Closeout Provide technical results and statistics of the migration to the Migration Manager during Closeout
Technical Migration Team (TMT)[29] (May include Data Architects, Business Analysts, Database Administrators, Programmers, Technical Writers, & and experts in various technology technologies employed during migration)	Contribute technical expertise during Planning Document legacy data architecture during Planning, if necessary Analyze & and Design Migration Data Architecture (staging & target) Execute the Data Migration Plan Develop migration procedures and perform trial migrations; refine procedures as needed Validate data migration results Compile technical results of migration on behalf of the Migration Technical Lead

2.2.8 Data Migration Planning Deliverables

Artifacts of the planning phase include:

- A dedicated *Project Management Plan*
- A *Data Migration Plan*
- A *Data Migration Requirements* document

[29] *2006 Best Practices for Data Migration,* page 6

- A *Risk Mitigation Plan*
- A *Risk Mitigation Matrix*
- A *Consolidated Legacy Entity Relationship Diagram (ERD)*
- A *Communications Plan*
- A *Data Conversion Plan*
- A *Data Integration Plan*
- A *Parallel Operation Plan*
- A *Data Migration Quality Plan*
- A *Data Migration Data Quality Remediation Plan*
- A *Data Migration Archival Strategy*
- A *Data Migration Test Plan*
- A *Data Migration Roles & Responsibilities* document

2.2.9 Data Migration Planning Checklist

The MMT may use the checklist provided in Appendix C to ensure that all aspects of Planning set forth in this methodology are accounted for in the *Data Migration Plan*. Once the plan is submitted to the EDM Team, the same checklist may be used, with supporting commentary for feedback, to evaluate the thoroughness of the plan.

2.3 Data Migration Analysis and Design

2.3.1 Analysis and Design Overview

Accurate design communicates the structural and procedural data requirements to the development team. The steps of Planning and Analysis & Design are logically inter-related, as many of the Planning activities depend upon results from Analysis and Design activities. The sequence of Design activities is not critical because much of the work might be performed concurrently and iteratively. Inter-dependency among activities will require a sequence when necessary. While not every Analysis and Design activity might occur for every migration, a thorough *Data Migration Plan* shall include either requirements for the Analysis and Design activities or justifications for not including a particular activity.

From the Legacy (or source) Data Architecture, the completed analysis, and the *Data Migration Plan*, the TMT must design the Migration Data Architecture. This architecture is comprised of:

- The 'as is' legacy architecture (blueprint)
- The 'to be' Staging Area Data Architecture, if necessary
- The 'to be' Target Data Architecture
- The correlations ('mappings') showing inter-architecture relationships and instructions to turn data from one data architecture into valid data in the next (source into staging or target, staging into target)

The data structures and data migration procedures must satisfy the data migration requirements. The resulting artifacts consist of fully-attributed logical data models, data dictionaries, function (or process) models, mappings between "as is" and "to be " data structures with corresponding business rules and transformation logic, and any other applicable artifact defined in the *Federal Student Aid Integrated Architecture Framework*. These artifacts should then be vetted through the EDM Team and business area stakeholders to ensure that:

- Only the data required to sustain the future application is migrated, and an audit trail is provided back to the legacy system data
- The design ensures that all data in scope is migrated accurately and completely
- The design ensures that the integrity of the migrated data is maintained
- The design minimizes disruption to the business during transition
- Mappings provide for a detailed inventory of all data and systems architecture, as well as identification of any data issues relevant to the conversion during the early phases of the project

2.3.2 Data Migration Analysis and Design Tasks and Subtasks

The following table presents the four major tasks of the Data Migration Analysis & Design Phase. Each task is broken down further in subtasks. Subtasks may occur in parallel or in the sequence shown in this chart. The Data Migration Project Manager, in collaboration with the DMT, will determine the order in which these tasks will be performed.

Table 7: Data migration analysis and design.

Data Migration Analysis and Design Phase			
Perform Data Migration Analysis	**Determine Data Security Controls**[30]	**Design Data Migration Environments**	**Design Data Migration Procedures**
Analyze Current Environment	Determine Enterprise Management and Operational Security Controls	Design Staging Area	Design Data Staging Procedures
Evaluate Data Migration Technology		Design Target Data Architecture	Design Data Cleansing Procedures
Evaluate Data Quality		Correlate Migration Data (Source/Staging/Target)	Design Data Conversion Procedures
Perform Data Profiling		Determine Data Migration Technology Configuration	Design Target Data Migration Procedures
Critical Success Factors for Data Migration Analysis and Design			Design Data Validation Procedures
			Design Data Quality Remediation Procedures.
			Refine Data Migration Test Plan
Disseminate Analysis & Design Artifacts			
Data Migration Analysis & Design Checklist			

[30] *IPM Data Management Plan*

2.3.3 Perform Data Migration Analysis

2.3.3.1 Analyze Current Environment

The MMT, and specifically the Technical Lead, must review all compiled artifacts related to the legacy environment to fully understand the IT environment in place and any constraints or technical limitations. It is important to determine whether these limitations and/or constraints will remain in the new target environment or not (e.g. move from a mainframe solution to a client-server environment; data validation performed at the front end through the Graphical User Interface (GUI) or on the back end at the database level.) This information will affect the design of the data migration activities and procedures.

2.3.3.2 Evaluate Data Migration Technology

The MMT, and specifically the Technical Lead, must review all available technology options. Should the technology in place not meet the requirements, the team should select and recommend tools best suited – within available funding – to satisfying the requirements of the data migration.

Data Migration infrastructure should be in place during the design phase of the project.

2.3.3.3 Evaluate Data Quality

The MMT, and specifically the Data Steward, must evaluate the current state of the legacy data in accordance with the *Data Migration Plan*, and specifically the *Data Quality Metrics*.

Data Quality[31] refers to the reliability and effectiveness of data. Data are of high quality "if they are fit for their intended uses in operations, decision making and planning" (J.M. Juran). Alternatively, data is determined to be of high quality if it correctly represents the real-world construct to which it refers.

The MMT, and specifically the Data Steward, determines how to assess the quality of the legacy (source) data (*Data Quality Metrics*) and how to remedy identified quality issues in collaboration with the Business Owners and Stakeholders. The actual remediation steps will be defined after the data analysis is performed, and after and the Data Quality Metrics are applied. These actions result in the refined *Data Quality Remediation Plan*.

There are two types of potential data quality issues, each of which requires a different set of remediation activities:

- Non-compliance with the metadata structure (social security number data field containing letters instead of numbers; state abbreviations that are non-compliant with the official state abbreviations (such as “ML” instead of “MD” for Maryland).
- Bad or corrupted data in the database (incomplete social security number; multiple records for the “same” person, mandatory data fields left blank)

[31] Wikipedia: http://en.wikipedia.org/wiki/Data_quality

The procedures for correcting any data quality issues shall be developed and executed during the Implementation stage of the migration. Data Quality measures the compliance with business and data validation rules implemented by the analyzed application. Additional analysis can be done by spot-checking with custom-build SQL-queries for bad data such as corrupted records. The results of the data analysis determine the need for data cleansing, and the data quality metrics will indicate what specific action needs to be taken.

The outcome and the agreed-upon remediation activities may have a significant effect on the timeline of the data migration project. Often, the responsibility for deciding on how and when to repair the identified issues lies outside the immediate project, and becomes a dependency that poses a potential risk to the project.

2.3.3.4 Perform Data Profiling

Data Profiling[32] is the initial assessment of the legacy data (structure and content) to understand and determine any quality challenges. There are two types of Data Profiling:

- Metadata profiling: assessment and examination of the data structures in place. The results from this activity can be used to evaluate compliance with enterprise-wide standards.
- Content profiling: assessment and examination of the data content. The results of this assessment reflect the quality of the content of the data captured, and identify issues that will be resolved through data cleansing.

The assessment is a process whereby the team examines the data available in an existing database and collects statistics and information about that data. The purpose of these statistics is to:

- Give metrics on data quality, including whether the data conforms to company standards
- Assess the risk involved in integrating data for new applications, including the challenges of joins
- Monitor and track data quality
- Assess whether metadata accurately describes the actual values in the source database
- Assess the risk involved in integrating data for new applications, including the challenges of joins

Understanding data challenges early in any data migration project will reduce surprises later on in the project. Finding data problems late in the project can incur time delays and project cost overruns.

[32] Wikipedia: http://en.wikipedia.org/wiki/Data_profiling

Profiling activities should follow these three steps in the order:

- **Step 1 — Column Profiling:** Provides critical metadata.
- **Step 2 — Dependency Profiling:** Identifies intra-table dependencies. Dependency profiling relates to the normalization of a data source, and addresses whether or not there are non-key attributes that determine or are dependent on other non-key attributes. The existence of transitive dependencies here might be evidence of second-normal form.
- **Step 3 — Redundancy Profiling:** Identifies overlapping values between tables. This is typically used to identify candidate foreign keys within tables, to validate attributes that should be foreign keys (but that might not have constraints to enforce integrity), and to identify other areas of data redundancy. Example: redundancy analysis could provide the analyst with the fact that 80% of the time, the ZIP field in table A contained the same values as the ZIP_CODE field in table B.

Column profiling provides critical metadata, which is required in order to perform dependency profiling, and as such must be executed before dependency profiling. Similarly, dependency profiling must be performed before redundancy profiling.

The use of automated data profiling tools is a Best Practice in the data profiling step in data migration[33].

This step partially overlaps with blueprinting the state of the legacy architecture. Automated profiling tools might be used to facilitate the procedures, but the procedures might also be done manually in the absence of automated software. Once completed, a Data Profile contributes to Data Conversion and Data Quality Remediation. The documented results of the analysis become a resource for the design of the staging and target architectures. The *Data Profile Assessment* needs to be distributed and presented for approval as outlined in the Communications Plan to discuss and determine the criticality of the findings. The Data Quality Remediation effort reflects on this task in greater detail.

In addition, Federal Student Aid must analyze which of the identified Data Quality issues can be resolved through Data Cleansing. The questions below help in determine the most appropriate Data Cleansing approach and responsibilities:

a) Where do the identified data issues originate?

   There are two possibilities: They were introduced by Federal Student Aid applications or through data received from their business partners as part of the data exchange. Data issues introduced by non-Federal Student Aid sources may not be easily repaired because of their external ownership.

b) How can the identified data issues be repaired?

   It might be possible to implement strong validation rules at the front end (GUI) and prevent the entry of invalid data at the point of data entry; or validation rules can be implemented at the database level.

[33] *A Strategic Approach to Data Migration,* page 2

c) Do the data issues refer to historical data only (e.g. data older than 5-10 years)?

It is possible that improved business and validation rules have been implemented after detection of these data issues, and that newer data is in better condition. A decision needs to be made as to whether it is worthwhile and/or necessary to repair the historical data records.

d) Are the identified data issues caused by missing validation rules?

Design and implementation of proper business rules could repair these issues.

e) Who owns the data and is responsible for the data quality?

Usually, the business owners are responsible to ensure high data quality from a business perspective. The business owners, in collaboration with the Data Steward(s), should define the necessary steps for data cleansing and prepare a plan and timeline for implementation.

2.3.3.5 Critical Success Factors for Data Migration Analysis & Design

Best practices recommend the following critical success factors be set as goals while designing target data structures and data migration procedures:

- Understand data requirements (architecture and business rules),
- Design comprehensive data migration procedures upon understanding of data content, quality, and structure[34], and
- Leverage standardized data structures (Enterprise Data Architecture).

2.3.4 Determine Data Security Controls[35]

2.3.4.1 Determine Enterprise Management and Operational Security Controls[36]

Each data migration project needs to operate within the security controls in place at Federal Student Aid. It is important to understand these requirements in to ensure proper handling of these data assets. Therefore, the MMT must analyze and follow the processes outlined in the "*Handbook for Information Assurance Security Policy Information Assurance Program March 31, 2006*" for protection of all data at each source and during the migration of the data between sources. This guide also addresses the following operational security controls of:

- Personnel security
- Physical and environmental protection
- Production controls
- Contingency planning

---

[34] *Strategic Approach to Data Migration,* page 3
[35] *IPM Data Management Plan*
[36] *IPM Data Management Plan*

- System hardware controls
- Software maintenance controls
- Data integrity/validation controls
- Documentation standards
- Security awareness training
- Incident reporting
- Data encryption

.

2.3.5 Design Data Migration Environments

2.3.5.1 Design Staging Area

The data migration transfers data from one or more legacy (source) data stores to a target data store. The migration is seldom a *direct* transfer between the source and target. Staging the data in an interim location allows for additional processing. Most of the activities created during planning are best performed in a staging area, specifically Data Conversion / Transformation and Data Quality Remediation.

It is possible (and at times preferable) to perform quality remediation (fixing errors) within the actual legacy (source) data location, but it is *not* recommended to perform quality remediation within the target data location, especially if the target is already an operational data location. Some commercial data conversion / transformation software packages perform the data changes within a proprietary data area within the tool. This is in itself a staging area of sorts, but is not always within the control of the migration team.

If either of these activities requires a custom data area, then the TMT must design the Staging Area Data Architecture. The data structures must house the interim data during the (trial) migration, and the procedures must populate the structures throughout the migration(s). To ensure the success and validity of the data migration procedures and data architecture, it is essential that the staging environment mirror the production environment. Otherwise, the results of the trial migrations might not be representative.

Another strong reason for processing the migrating data within a staging area is the need to reconcile or integrate data coming from separate data sources. If data representing the same information is structured differently in different sources, a staging area provides the means to consolidate all of the source data and validate it before transferring the data to the target data location. Also, if data is being migrated in phases — different sources at different times or only a part of the data at a time — the staging area provides an area to process and transform the migrating data before loading it to the target data location.

The proposed Staging Data Architecture needs to be distributed and presented for approval as outlined in the *Communications Plan*.

2.3.5.2 Design Target Data Architecture

The TMT must establish the Target Data Architecture. The target data structures must house the data to support the target application, and the procedures must populate the target data store from the staging area (or the source data store, if a staging area is not used during the migration). The

data structures to house the target data should satisfy the data requirements of the target application or business area.

In addition to the data structures required for the target data store, the TMT must design procedures to:

- Move either the staged data or the source data into the target data store
- Integrate migrated data into the target data store, if the target data store is already populated
- Validate that previously existing data structures and functionality remain intact after integrating migrated data into a target data store with already operational data. This scenario also occurs when performing sequential data migrations due to large volumes into one target data store
- Validate the migrated data for completeness and accuracy

The proposed Target Data Architecture needs to be distributed and presented for approval as outlined in the *Communications Plan*.

2.3.5.3 Correlate Migration Data (Source/Staging/Target)

The TMT must now correlate, or map, the individual architectures to each other, providing a roadmap for the data — and the migration team — to follow from the original data source to the final (target) data source. If an interim staging area is used, then the path should, obviously, pass through the staging area.

In addition, the correlations must also include any mapping of the data to the rules, translations, and/or transformation procedures to which the data must adhere when moving from one data location to another. The *Data Correlation Report* needs to be distributed and presented for approval as outlined in the *Communications Plan*.

2.3.5.4 Determine Data Migration Technology Configuration

Available resources, such as the architectural design tools and metadata repositories available at Federal Student Aid, and the migration method(s) selected for the effort, determine most of the technology of the data migration. Existing technology, such as the source data storage devices and software, must be captured as part of the Legacy (or 'Baseline'[37]) Data Architecture, which is simply the way the legacy data resources look and operate prior to the migration.

The MMT must determine all of the technical aspects of the staging area (if used), target environment, and actual movement as well as the validation of the data itself.

As part of this task, the Data Migration Team must select the technology to access the legacy data during the migration. For example, the technology may be any standard ETL software in use at Federal Student Aid, or may be custom programming[38].

[37] *A Practical Guide to Federal Enterprise Architecture, February 2001 (CIO Council),* page 5
[38] *The Complete Data Migration Methodology*, page 6

Questions to consider in deciding whether available migration software may be used include:

- Is the migration software compatible with both the source AND the target storage hardware?
- Does the migration software facilitate restoring the pre-migration state in the event of errors?
- Does the migration software support non-disruptive data migration (i.e. legacy application continues to operate during the migration) if this is a requirement of the migration?
- How quickly is data transfer performed?

If the software available at Federal Student Aid does not satisfy the above requirements, and custom software or another commercial package must be selected, additional considerations include:

- Is the migration software supported on the intended operating system?
- Is the migration software supported on the intended hardware?
- Does the migration software support differences in volume size between source and target data?
- Can the operational environment accommodate the processing requirements of the software?
- Does the migration software support online data migration?

If existing Federal Student Aid software satisfies the requirements of the data migration, then the TMT must configure the software to perform the actual migration at hand. If the existing software does *not* satisfy the requirements, then the team may take one of two actions: (1) seek approval from the business-area owners and stakeholders for acquisition of new software, or (2) determine requirements for custom procedures to satisfy the needs of the migration. The proposed Data Migration Technology Configuration, including all necessary decision points, must be distributed and presented for review and approval as outlined in the Communications Plan. This design can also be used a blueprint for future projects. However, because each data migration project has different requirements and needs, the design must be adjusted to support the needs of the particular project.

2.3.6 Design Data Migration Procedures

2.3.6.1 Design Data Staging Procedures

The TMT designs the necessary procedures to extract the data from the legacy (source) data store, transport the data to the staging area, and perform any necessary translations, or transformations, prior to populating the staging area. Sources of input for this task include:

- Blueprint of the existing data architecture
- Design of the staging area data store
- Data correlation report
- Any identified rules, translations, or transformations that the data must undergo to meet the staging/target data structure requirements.

In cases of data migration projects where a staging area is not used, the data extract and transport procedures are still required. This activity may overlap with the activity of designing target migration procedures. Next to the design of this task, and of all following design tasks, the team must also develop a set of test cases to validate the procedures.

2.3.6.2 Design Data Cleansing Procedures

The TMT designs the necessary procedures to fix errors and refine the source data, either within the legacy (source) data store or within the staging area, based on input from the business owners and other stakeholders. Sources of input for this task include:

- Data Quality Plan and metrics
- Data Profiling Assessment and Report
- Data Quality Remediation Plan
- Stakeholder decisions

As part of this activity, the ownership and responsibility for the execution of the data cleansing task must be determined. In many cases, when the TMT consists mainly of contracting staff, Federal Student Aid employees will perform the data cleansing. This situation creates a dependency for the project, as the staff performing the task may be outside of the immediate project team, and therefore not managed by the Data Migration Team Lead. This situation should be treated as a potential risk with respect to the timely completion of the task.

2.3.6.3 Design Data Conversion Procedures

The TMT designs the necessary procedures to convert the source data to the proper values and formats required in the staging and target data store. Sources of input for this task include:

- Blueprint of the existing data architecture
- Design of the staging area and the target data store
- Data correlation report
- Any identified rules, translations, or transformations that the data must undergo to meet the staging/target data structure requirements

2.3.6.4 Design Target Data Migration Procedures

The TMT designs the necessary procedures to extract the data from the staging or source data stores as appropriate, transport the data to the target data store, and in case no staging area is used, perform any necessary translations or transformations prior to populating the target data store.

2.3.6.5 Design Data Validation Procedures

The TMT designs the necessary procedures to validate the integrity of the data content at each stage of the migration. Validation procedures must also support validation of the completeness and accuracy of the migrated data.

2.3.6.6 Design Data Quality Remediation Procedures

The TMT designs the procedures to be used to remediate identified data quality issues through the data profiling. These procedures will be developed in close collaboration with Federal Student Aid business owners to meet current (and future) business requirements, and in collaboration with the EDM Team with regard to compliance with Federal Student Aid data standards.

2.3.6.7 Refine Data Migration Test Plan

All individual test cases must be consolidated into the overall *Data Migration Test Plan*. The TMT uses this detailed information to update and refine the original plan through sequencing the execution of these test cases and identification of any dependencies.

2.3.7 Data Migration Analysis & Design Deliverables

Artifacts of the planning phase include:

- A *Data Profiling Assessment & Report (Legacy System)*
- A *Data Migration Data Quality Remediation Procedures Design*
- A *Migration Data Architecture* covering:
    - *Staging Area* and *Target Data Architecture Design including*
        - *Logical Data Model (Staging and Target Area)*
        - *Physical Data Model (Staging and Target Area)*
        - *Data Dictionaries (Staging and Target Area)*
    - *Migration Data Correlation Report*
    - *Data Migration Technology Configuration*
- *A Data Migration Procedures Design*
    - A *Data Staging Procedures Design*
    - A *Data Cleansing Procedures Design*
    - A *Data Conversion Procedures Design*
    - A *Target Data Migration Procedures Design*
    - A *Target Data Validation Procedures Design*
- A *Data Migration Test Plan (refined)*
- A *Data Load Plan*
- A *Data Back-up & Recovery Plan (including back-out)*

2.3.8 Design Checklist

The MMT may use the checklist provided in Appendix C to ensure that all activities of Analysis and Design set forth in this methodology are accounted for in the *Data Migration Plan* and the *Migration Data Architecture*. As the *Data Migration Plan* and/or *Migration Data Architecture* are submitted to the stakeholders and EDM Team for approval, the same checklist may be used, with supporting commentary for feedback, to evaluate the proposed architecture and thoroughness of the documentation.

2.4 Implementation

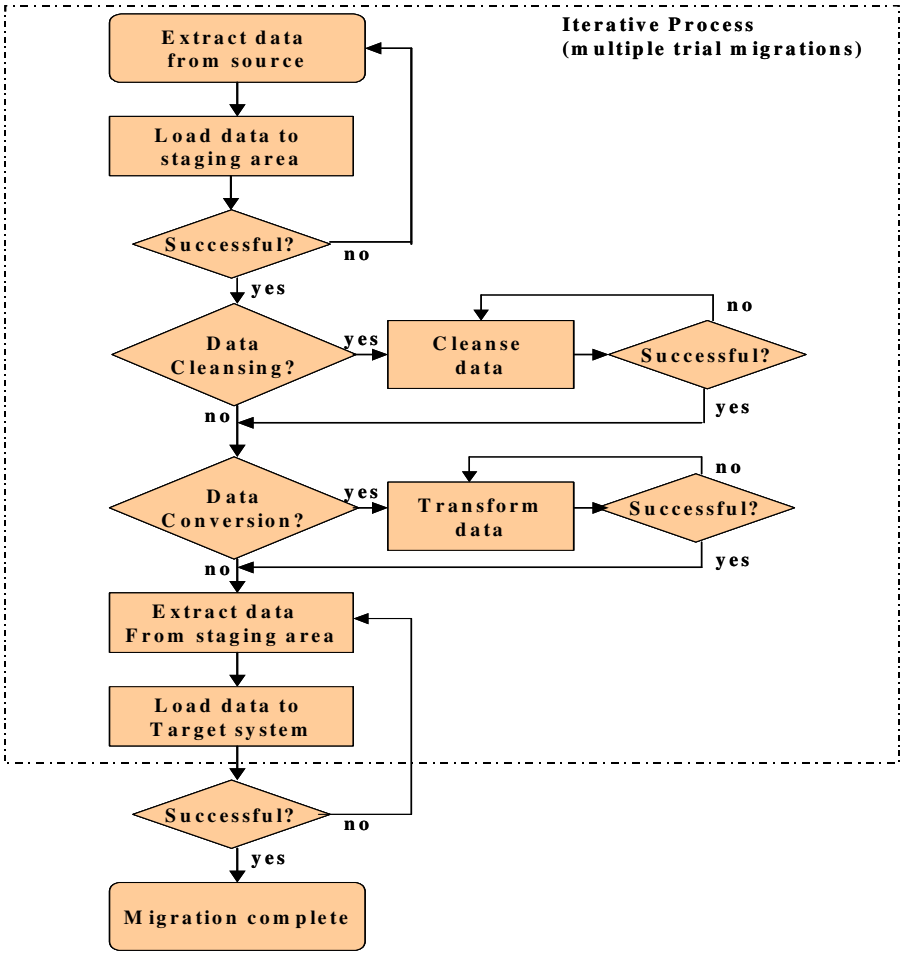
2.4.1 Implementation Overview

The steps of Implementation and Validation are logically interdependent, much like the steps of Planning and Analysis and Design. Although not every activity may occur for every migration, the ones that *do* occur should generally follow the sequence shown below. There is one exception: As discussed in several Best Practice articles, a decision should be made during Planning as to the most efficient and effective time to do Data Cleansing and Data Conversion. Depending on how passive or active these tasks are in a specific data migration, they might occur further upstream than shown in the *Data Migration Implementation Process Flow*.

If the data migration covers multiple source systems, repeat the steps of extracting the data and loading it in to the staging area until all source data has been loaded. The staging area supports the integration of all data. The *Data Migration Plan* determines the exact process for such integration.

Figure 6 shows the high-level process flow of the Implementation Phase.

Figure 6: Data Migration Implementation Process Flow.



2.4.2 Data Migration Implementation Tasks & Subtasks

The following table presents the six major tasks of the Data Migration Implementation Phase. Each task is broken down further in subtasks. Subtasks may occur in parallel or in the sequence shown in this chart. The Data Migration Project Manager, in collaboration with the DMT, will determine the order in which these tasks will be performed.

Table 8: Data migration implementation.

<table>
<tr><th colspan="6">Data Migration Implementation Phase</th></tr>
<tr><th>Develop Data Migration Procedures</th><th>Stage Data</th><th>Cleanse Data</th><th>Convert/ Transform Data</th><th>Migrate Data</th><th>Post-migration</th></tr>
<tr><td>Configure Resources</td><td>Create Staging Area</td><td>Cleanse Data according to Data Remediation Plan</td><td>Convert/ Transform Data</td><td>Perform Trial Migration</td><td>Operate Legacy and Target Environment in Parallel</td></tr>
<tr><td>Develop and Test Data Migration Procedures</td><td>Populate Staging Area</td><td>Validate Cleansed Data</td><td>Validate Converted/ Transformed Data</td><td>Validate Results of Trial Migration</td><td>Validate Parallel Operation</td></tr>
<tr><td>Develop and Test Validation Procedures</td><td>Integrate Staged Data</td><td rowspan="3" colspan="2"></td><td>Obtain Approval for Full Migration into Production Environment</td><td>Release Data Environments</td></tr>
<tr><td>Develop and Test Data Cleansing Procedures</td><td>Validate Staged Data</td><td>Perform Full Migration (Deployment)</td><td rowspan="2"></td></tr>
<tr><td>Develop and Test Data Conversion/ Transformation Procedures</td><td></td><td>Validate Results of Full Migration</td></tr>
<tr><td>Establish Access to Staging and Target Area</td><td colspan="5"></td></tr>
<tr><td>Critical Success Factors for Implementation</td><td colspan="5"></td></tr>
<tr><td colspan="6">Data Migration Implementation Artifacts</td></tr>
<tr><td colspan="6">Data Migration Implementation Checklist</td></tr>
</table>

Using the *Data Migration Data Quality Plan* (here, using the Data Loss Tolerance and Data Quality Metrics as a benchmark), the TMT must execute the validation procedures to measure the success of each stage of migration, and must also validate any stage of the implementation

that changes the data (such as format, content, or location). Any time the resulting data fails the validation, the procedures of that migration stage should reverse, or "roll back," the data and make any needed corrections. The trial migrations should be repeated until acceptance and/or success criteria are met and the migration procedures and environments are ready for deployment.

2.4.3 Develop Data Migration Procedures

2.4.3.1 Configure Resources

In some cases, data migration planning and implementation are separately-funded tasks. This funding arrangement can put constraints on resource availability. If not done during project start-up, all resources to be used during the migration must be acquired, assigned, installed, and/or configured prior to implementation.

Do the following:

- Assign personnel of the TMT to the various tasks required to migrate the data
- Define Create, Read, Update and Delete (CRUD) Matrix outline access privileges to Staging Area and Target Data Store for team members
- Request or acquire the data access required by the technical team to implement the various procedures (if not already done)

The following technical steps must now occur:

- Establish test environments
- Make test environments available to the technical team
- Install and/or configure all software involved in the migration[39]

These steps carry a potential risk and dependency because both the creation of the test environment and the configuration and implementation of the software are most likely performed by staff outside the immediate project. Therefore, those staff members and their work are not under the direct control of the Project Manager.

2.4.3.2 Develop and Test Data Migration Procedures

The TMT develops and tests the data migration procedures in accordance with the *Data Migration Plan,* the *Data Migration Test Plan,* and the documented *Migration Data Architecture*.

2.4.3.3 Develop and Test Data Validation Procedures

The TMT develops and tests the data validation procedures in accordance with the *Data Migration Plan*, the *Data Migration Test Plan*, and the *Migration Data Architecture*; as well as

[39] *Microsoft CRM Data Migration Framework,* page 8

in compliance with all applicable business rules and processes. These procedures must be capable of confirming that data loss and accuracy is within allowable parameters during each stage of the migration. If 100% migration of data between extremely different environments is not feasible, the MMT must consult with business stakeholders and revise tolerances for data loss and update the data loss tolerance in the *Data Migration Plan* with the revised tolerances. The validation procedures at each stage must take these tolerances into account.

The procedures for validating the data at each stage of the migration may not be exactly the same, and not every migration effort will perform the same stages. However, the closer the procedures at each stage are to achieving uniformity, the easier it will be for the technical team to follow the data back through the process, should an error occur during the migration.

2.4.3.4 Develop and Test Data Cleansing Procedures

The TMT develops and tests the data cleansing procedures in accordance with the *Data Migration Plan*, the *Data Migration Test Plan*, and the *Data Migration Plan*, specifically the *Data Quality Metrics* and *Data Loss Tolerance*. If the procedures designed for data cleansing cannot be scripted or saved, possibly because they are steps to follow in a commercial software package or require significant manual operation, then they should be thoroughly documented and tested at this stage.

2.4.3.5 Develop and Test Data Conversion/Transformation Procedures

The TMT develops the data conversion and/or transformation procedures in accordance with the *Data Migration Plan,* the *Data Migration Test Plan,* and the *Data Migration Plan*, specifically the *Data Conversion Plan*. As mentioned before, if the procedures designed for data conversion cannot be scripted or saved, possibly because they are steps to follow in a commercial software package or require significant manual operation, then they should be thoroughly documented and tested at this stage.

2.4.3.6 Establish Access to Staging and Target Area

The TMT must coordinate with the proper points of contact at Federal Student Aid to gain all necessary access to the software and establish the user accounts that are required to perform all operations of the data migration. Omitting or delaying this task can result in a delay of the data migration project.

This task carries also a potential risk and dependency because granting access to the staging and target environment software and establishing user accounts will be performed by Federal Student Aid staff outside of the immediate project and are, therefore, not managed by the Project Manager.

2.4.3.7 Critical Success Factors for Implementation

Best Practices recommend the following critical success factors during migration of data from one or multiple source data stores to a target data store:

- Execution of a thorough and detailed *Migration Test Plan*

- Expertise in all data migration software[40]
- Expertise in all COTS and custom software used during the migration
- Comprehensive understanding of *Data Migration Plan* and *Migration Data Architecture*[41]
- Following and compliance with *Risk Mitigation Plan*
- Following and compliance with change control and quality assurance procedures

2.4.4 Stage Data

2.4.4.1 Create Staging Area

The data staging area can be created manually and/or automated. Either the TMT or Federal Student Aid staff can perform this critical task. Completion of this task is a precondition for using the staging area. Missing or delaying this task can cause a delay of the migration project.

This task carries also a potential risk and dependency as the creation of the test environment and the configuration and implementation of the software are most likely performed by staff outside of the immediate project and are, therefore, not managed by the Project Manager.

2.4.4.2 Populate Staging Area

The TMT executes the data migration procedures that transport (and transform, if applicable) data from the source data store to the staging area in accordance with the *Data Migration Plan* and documented *Migration Data Architecture*. If the staging of the data is to be done in stages (often due to multiple sources), then these steps must be performed according to the established schedule.

After completion of the task, the team may collect statistics for reporting and validation purposes, as laid out in the *Data Migration Plan*.

2.4.4.3 Integrate Staged Data

This step may not be independent of the steps required to populate the data in the staging area. The importance of *integrating* data from multiple sources must be clearly addressed in any migration plan. Depending on the age and condition of the source data, it might even be necessary to integrate multiple renditions of the same data from a single source, an activity that is commonly referred to as *reconciling* the source data.

2.4.4.4 Validate Staged Data

The TMT executes the validation procedures to measure the success of the staging of the source data. Upon completion, the team should then compare the results to the *Data Quality Metrics* and acceptance criteria set by the stakeholders and disseminate the outcome according to the

[40] *Microsoft CRM Data Migration Framework,* page 8
[41] *Strategic Approach to Data Migration,* page 3

Communications Plan. Also, if the threshold cannot be met after several trial migrations, stakeholders need to approve or disapprove moving forward, a decision point that is a dependency and risk to the project.

When multiple source systems are being merged into one target system, it is recommended to perform data validation after each load from each source system.

2.4.5 Cleanse Data

2.4.5.1 Cleanse Data according to Data Remediation Plan

The TMT must execute any procedures defined for measuring data quality. The data quality may be evaluated in the source data store, in the staging area, or in both. The TMT should then evaluate the results in the context of the *Data Quality Metrics* from the *Data Migration Plan*. If data errors need to be corrected, the *Data Quality Remediation* should be followed.

This step carries both potential risk and dependency, as the data cleansing is most likely performed by Federal Student Aid staff outside of the immediate project that are not directly managed by the Project Manager.

2.4.5.2 Validate Cleansed Data

The TMT must execute the validation procedures to measure the success of the cleansing stage of the data migration.

2.4.6 Convert/Transform Data

2.4.6.1 Convert/Transform Data

The TMT must execute any procedures defined for converting the source data format to the format required by the staging and/or target data store. If the data content must undergo transformation, the TMT should execute those procedures during this stage of the migration.

2.4.6.2 Validate Converted/Transformed Data

The TMT must execute the validation procedures to measure the success of the Data Conversion and/or Transformation. With commercial ETL software, these procedures may be built into the actual execution of the transformation. The resulting statistics should then be collected and presented for review.

2.4.7 Migrate Data

2.4.7.1 Perform Trial Data Migration

The TMT must execute the data migration procedures that transport (and transform, if applicable) data from the staging area, or the source data store, to the target data store in accordance with the *Data Migration Plan* and the documented *Migration Data Architecture*. These procedures may be performed on the full set of data or, if planned, on a representative subset. If the target environment is already operational, the target of the trial migration may be a test environment that mirrors the production environment. The results of the trial data migration must be validated immediately upon completion.

If the validation deems the migration 100% successful, and the trial migration was done on the full set of source data, the results may be accepted as the full migration as soon as the business area stakeholders approve them. If errors occur during the trial migration, or if the migration was done on an incomplete set of data or in a test environment, the necessary corrections must be made and the trial migration must be repeated.

If the trial migration was successful on a subset of data or on data directed to a test environment, the team may proceed to do the full data migration.

2.4.7.2 Validate Results of Trial Migration

The TMT must execute the validation procedures to measure the success of the trial migration and collect the results for review.

2.4.7.3 Obtain Approval for Full Migration into Production Environment

Once a successful trial migration has been executed and the results have been reviewed and approved by the business area stakeholders and the EDM Team, the TMT must proceed with the execution of the full data migration. If the final, successful trial migration satisfied all requirements and success criteria, the migration may be released for production.

2.4.7.4 Perform Full Data Migration -- Deployment

Upon approval, the TMT will execute the data migration procedures in the production environment in accordance with the *Data Migration Plan,* and documented *Migration Data Architecture*.

If resource limitations or the migration of large volumes of data make it necessary, the full data migration may be done in stages.

2.4.7.5 Validate Results of Full Migration

The TMT must execute the validation procedures to measure the success of the full data migration and collect the results for review.

2.4.8 Post-Migration

2.4.8.1 Operate Legacy & Target Environments in Parallel

Once the full migration has been executed, validated, and approved, the business area stakeholders may desire that the legacy system and the target system operate in parallel for a period of time. This decision, the duration of parallel operation, and the criteria for concluding the parallel operation should be captured in the *Data Migration Plan,* specifically in the *Parallel Operation Plan.*

2.4.8.2 Validate Parallel Operation

A "trial" period of parallel operation may be conducted to affirm that the target system is operating correctly. If errors are detected in the operation of the target system during the parallel operation period, then (1) the legacy system may continue to operate according to original requirements, and (2) the data migration may be revisited, repaired, and repeated.

2.4.8.3 Release Data Environments

Upon confirmation that the target environment is operating in a satisfactory manner, the source data environment may be released. If the source environment is to be retired from operations, retirement may occur at this time.

Any staging area, if used, would generally be taken out of operation at this point in accordance with the archival strategy developed as part of the overall *Data Migration Plan.* However, it is important that the environment not be purged entirely, in case errors in the data migration are discovered. It may be possible to restore the staging environment and re-populate the target environment without having to repeat the data migration procedures.

If the staging area is part of a long-term solution – such as an interim data store between an operational data store and a data warehouse – it should be approved and placed into full-time operation. The target environment should also be ready to begin operation as a full-time application data store.

2.4.9 Data Migration Implementation Artifacts

- Developed and fully tested
    - *Data Migration Procedures*
    - *Data Validation Procedures*
    - *Data Cleansing Procedures*
    - *Data Conversion/Transformation Procedures*
- An integrated, validated and documented *Staging Area*
- A validated and documented *Target Data Store*
- A *Data Cleansing Report*
- A *Data Conversion Report*
- *Trial Migration Results*
- *Full Data Migration Results*
- A *Parallel Operations Report*

2.4.10 Implementation Checklist

The MMT may use the checklist in Appendix C to ensure that all aspects of Implementation set forth in this framework are accounted for in the *Data Migration Plan* and followed during the migration. The EDM Team may use the same checklist, with supporting commentary for feedback, to evaluate the thoroughness and completeness of the data migration.

2.5 Data Migration Closeout

2.5.1 Closeout Overview

Once the operational scope of the data migration project is finished, the final data migration documentation must be prepared and submitted to the business area stakeholders and the EDM Team. The steps in closing out a data migration project are:

- Document data migration results.
- Document lessons learned.
- Perform knowledge transfer.
- Communicate data migration results and lessons learned.

2.5.2 Data Migration Closeout Tasks & Subtasks

The following table presents the four major tasks of the Data Migration Closeout Phase. Each task is broken down further in subtasks. Subtasks might occur in parallel, or in the sequence shown in this chart. The Data Migration Project Manager, in collaboration with the DMT, will determine the order in which these tasks will be performed.

Table 9: Data migration closeout.

Data Migration Project Closeout Phase			
Document Data Migration Results	**Document Lessons Learned**[42]	**Perform Knowledge Transfer**	**Communicate Data Migration Results**
Document Data Migration Results	Document Lessons Learned	Perform Knowledge Transfer	Disseminate Data Migration Results
Critical Success Factors for Data Migration Closeout		Provide Stakeholder Training	Obtain Approval for Project Completion
Disseminate Closeout Artifacts			
Data Migration Closeout Checklist			

2.5.3 *Document Data* Migration Results

The MMT, in tandem with the TMT, shall compile statistics from the data migration.[43]

2.5.3.1 Critical Success Factors for Data Migration Closeout

Best Practices dictate that the following critical success factors be achieved while summarizing data migration results:

- Understand Data Requirements

[42] *IPM Data Management Plan*
[43] *2006 Best Practices for Data Migration,* page 8

- Ensure that all user expectations are addressed and/or satisfied by the reported results
- Establish and/or follow a standard set of report formats for disseminating migration results
- Follow standardized and/or mandated project closeout procedures
- Obtain final stakeholder approval

2.5.4 Document Data Migration Lessons Learned

The MMT, in tandem with the TMT, shall document lessons learned from the data migration.[44]

2.5.5 Perform Knowledge Transfer

All plans, architectures, and results should be packaged and provided to the business area stakeholders as an audit trail of the data, and as a demonstrated example of a successful migration. This information helps make the data migration a repeatable process that future migration efforts can leverage.

2.5.6 Perform Stakeholder Training

Any stakeholder training in data migration procedures and/or the staging and target data stores may occur at this point. This training is critical to ensure a smooth transition of the newly developed system to Federal Student Aid.

2.5.7 Disseminate Data Migration Results

As with the plans and designs, it is crucial to communicate the results of the migration. The results and lessons learned shall be provided to the EDM Team to evaluate the methodology employed in the migration, and to determine whether revisions to the methodology are necessary to improve future data migration efforts. The results shall also be provided to the business area stakeholders, who shall determine whether the business requirements of the migration have been met. Once the data migration results have been reviewed and approved by the business area stakeholders and the EDM Team, all the documentation and artifacts should be made available to all other Federal Student Aid projects for analysis and leveraging of plans, procedures, and designs.

2.5.8 Obtain Approval for Project Completion

The approval of the data migration results by the business area stakeholders and the EDM Team is also considered the approval for project completion. The Data Migration Project Manager and the EDM Team close out the project.

2.5.9 Data Migration Closeout Artifacts

- A detailed and interpreted *Statistics Report* on Full Data Migration
- A *Lessons Learned Report*

[44] *2006 Best Practices for Data Migration,* page 8

- A *Training Plan* for Staging and Target Environment Configuration
- A *complete set of Data Migration Documents and Artifacts* in electronic and/or paper format.

2.5.10 Data Migration Closeout Checklist

The MMT may use the checklist in Appendix C to ensure that all aspects of data migration closeout set forth in this methodology are accounted for in the *Data Migration Plan* and in any data migration summary reports. Once results of the migration are submitted to the EDM Team, the same checklist may be used, with supporting commentary for feedback, to evaluate the thoroughness of the reported results.

Appendix A: Glossary

The following terms are used herein or are pertinent to content included herein:

Best Practice[45]: A management idea asserting a technique, method, process, activity, incentive or reward as more effective at delivering a particular outcome than any other technique, method, process, etc.

Column: A set of data values of the same type collected and stored in the rows of a table.

Database: A set of table spaces and index spaces.

Data Conversion[46]: The [transition] of one form of computer data to another.

Data Element: A generic term for an entity/class, table, attribute, or column in a conceptual, logical, and physical data model.

Data Migration[47]: The transferring of data between storage types, formats, or computer systems.

Enterprise Conceptual Data Model (ECDM): One of the initial components of Enterprise Data Architecture. The first enterprise level data model developed. The ECDM identifies groupings of data important to Lines of Business, Conceptual Entities, and defines their general relationships. The ECDM provides a picture of the data the enterprise needs to conduct its business. (**Reference**: *U.S. Department of Education Enterprise Data Architecture – Enterprise Data Standards and Roadmaps.)*

Enterprise Data Dictionary (EDD): One of the initial components of Enterprise Data Architecture. The EDD lists metadata objects and a complete description of the object at a sufficient level of detail to ensure that they are discrete and clearly understood. Such descriptions shall include, at a minimum, labels (names, titles, etc.) and definitions (or text descriptions), but may include additional descriptive metadata such as object type, classifications, content data type, rules (business, validation, etc.), valid and default values, etc. The EDD is the definitive source for the meaning of metadata objects. (**Reference:** *FSA-EDM*)

Enterprise Logical Data Model (ELDM): A component of a maturing Enterprise Data Architecture. The second enterprise level data model developed. It is the result of merging application level data model information into the existing Enterprise Conceptual Data Model (ECDM). The ELDM extends the ECDM level of detail. (**Reference**: *U.S. Department of Education Enterprise Data Architecture – Enterprise Data Standards and Roadmaps)*

eXtensible Markup Language (XML): A meta-markup language for describing data elements that is extensible because it does not have a fixed set of tags and elements.

eXtensible Stylesheet Language (XSL): A standard from the W3C for describing a style sheet for XML documents.

---

[45] Derived from Wikipedia "Best Practice" (http://en.wikipedia.org/wiki/Best_practice)
[46] Derived from Wikipedia "Data Conversion" (http://en.wikipedia.org/wiki/Data_conversion)
[47] Derived from Wikipedia "Data Migration" (http://en.wikipedia.org/wiki/Data_Migration)

Enterprise Data Standards and Roadmaps (EDSG): A component of a maturing Enterprise Data Architecture; rules and recommendations for the creation and updating of metadata objects and structures as well as for creating conceptual and physical models and schemas at both the enterprise and application level. (**Reference:** *FSA-EDM*)

Management Idea[48]: (a.k.a "Management fad") A change in philosophy or operations that sweeps through businesses and institutions, and then disappears when enthusiasm for it wanes.

Operational Data Store (ODS): An operational data store (ODS) is a type of database often used as an interim area for a data warehouse. Unlike a data warehouse, which contains static data, the contents of the ODS are updated through the course of business operations. An ODS is designed to quickly perform relatively simple queries on small amounts of data (such as finding the status of a customer order), rather than the complex queries on large amounts of data typical of the data warehouse

Schema (XML): A definition, written in eXtensible Markup Language (XML) syntax, of constraints for the content type and data type of XML tags.

Schema (Data): Any diagram or textual description of a structure for representing data. (**Reference:** *FSA-EDM*)

Table: A set of related columns and rows in a relational database.

Table Space: A portion of a database reserved for where a table will go. Table structure is the mapping of tables into table spaces.

Tag (XML): The markup portion of an Extensible Markup Language (XML) element surrounding the character data. The name of the tag reflects the content inside the XML element.

Target Data Store: The data store where the migrated and/or transformed data will be moved.

Uniform Resource Identifier (URI): The addressing technology for identifying resources on the Internet or a private intranet.

Uniform Resource Locator (URL): The address that defines the route to a file on a Web server.

Uniform Resource Name (URN): A name that identifies a resource on the Internet. Unlike URLs, which use network addresses (domain, directory path, file name), URNs use regular words that are protocol and location independent.

Valid (XML): A well-formed eXtensible Markup Language (XML) document that also matches the Document Type Definition (DTD).

Well-formed (XML): An eXtensible Markup Language (XML) document that has sufficiently specific grammar to be read and understood by an XML parser.

World Wide Web Consortium (W3C): An international industry consortium founded in 1994 to develop standards for the Web. The W3C has standardized many of the fundamental technologies of the Web, including HTML and XML, URLs and URIs, the SOAP protocol and the P3P privacy description.

[48] Derived from Wikipedia "Management fad" (http://en.wikipedia.org/wiki/Management_fad)

Appendix B: Abbreviations and Acronyms

The following abbreviations and acronyms are used herein or are pertinent to content included herein:

Abbreviation / Acronym	Applicable Term
CDM	Conceptual Data Model
ECDM	Enterprise Conceptual Data Model
ED	Department of Education
EDD	Enterprise Data Dictionary
EDM	Enterprise Data Management
EDMMG	Enterprise Data Management Master Glossary
ELDM	Enterprise Logical Data Model
FEA	Federal Enterprise Architecture
FEAF	Federal Enterprise Architecture Framework
FIPS	Federal Information Processing Standards
IT	Information Technology
ITSS	Information Technology System Services
LDM	Logical Data Model
OS	Operational System
PDM	Physical Data Model
PESC	Postsecondary Electronic Standards Council
XML	eXtensible Markup Language

Appendix C: Data Migration Project Review Checklist

Data Migration Review Checklist	Status			
	P	C	A	N/A
Overall				
Objective Described (data condition to be resolved)				
Business Value of Data Explained				
Included Artifacts Clear and Accurate				
Benefits				
Overall				
Benefits of hardware choices				
Benefits of software choices				
Benefits of Data Cleansing				
Benefits of Data Conversion				
Other				
Planning				
Project Planning				
Initiate Data Migration Project				
Establish Scope				
Establish Data Migration Roles & Responsibilities				
Develop Change Management Policies & Procedures				
Identify Critical Success Factors for Planning Phase				
Identify Risks / Constraints / Dependencies / Assumptions				
Funding				
Expertise Availability				
Personnel Availability				
Allowable Downtime				
Scheduling Constraints (Time)				
Legacy Environment Complexity				
Allowable Downtime				
Hardware Availability				
Physical Storage Incompatibility				
Application Performance Remediation				
Tolerance and Remediation of Planned Resources				
Tolerance and Remediation of Missed Requirements				
Other				
Develop Risk Mitigation Strategy				
Migration Requirements				
Determine Design Requirements				
Required Design Artifacts				
Migration Execution Performance Metrics				
Migration Execution Requirements				
Other				
Determine Time Requirements				
Determine Technology				
Data Storage Distribution				
Physical re-location Requirements				
Target Hardware Configuration				

Data Migration Review Checklist	Status			
	P	C	A	N/A
Target Software Configuration				
Homogeneous vs. Heterogeneous Storage				
Multi-vendor Storage Environment				
Target Data Capacity				
Other				
Determine Stakeholder Requirements				
Determine User Expectations				
Determine Data Security Requirements				
Source Data Protection (Recoverability)				
Access to Migrating Data				
Access to Migration Environment				
Access to Documentation				
Access to Legacy Environment				
Access to Interim Environment				
Access to Target Environment				
Other				
Consider Future Requirements				
Future Capacity Growth				
Durability of Target Data Storage				
Migratability of Target Data Storage				
Re-usability of Target Data Storage				
Scalability of Target Data Storage				
Other				
Current Environment				
Existing Data Related Artifacts				
Availability of Data Architecture Documents				
Blueprint Current Data Architecture				
Availability of Current Data Storage				
Profile Legacy Data				
Determine IT Infrastructure Requirements				
Develop Data Migration Plan				
Identify Technology Options				
Determine Data Migration Method				
Plan Data Content Management Strategy				
Data Conversion Plan				
Data Quality Metrics				
Data Loss Tolerance				
Data Quality Remediation Plan				
Data Integration / Reconciliation Plan				
Data Archival Strategy				
Other				
Plan Parallel Operation				
Plan Data Security Strategy				
Develop Data Migration Plan				
Analysis & Design				
Analysis				
Analyze Current Environment				
Evaluate Data Migration Technology				

Data Migration Review Checklist	Status			
	P	C	A	N/A
Evaluate Data Quality (Data Profiling)				
Correlate Data to Business Processes				
Identify Critical Success Factors for Analysis & Design Phase				
Determine Security Controls				
Design Enterprise Management Controls				
Design Operational Security Controls				
Access to Migration Environment				
Access to Documentation				
Design Technical Security Controls				
Access to Migrating Data				
Access to Legacy Environment				
Access to Interim Environment				
Access to Target Environment				
Design Data Environment				
Design Security Data Architecture				
Design Staging Area				
Design Target Data Architecture				
Correlate Migration Data (Source / Stage / Target)				
Legacy to Staging				
Legacy to Target				
Staging to Target				
Correlate Migration Data to Procedures				
Determine Technology Configuration (ETL, RDBMS, etc)				
Design Migration IT Infrastructure				
Design Migration Procedures				
Design Data Security Procedures				
Design Data Staging Procedures				
Design Data Cleansing Procedures				
Design Data Conversion Procedures				
Design Target Data Migration Procedures				
Design Data Validation Procedures				
Design Data Quality Remediation Procedures				
Refine Data Migration Test Plan				
Design Data Quality Reports and Process for Reconciliation				
Implementation				
Develop Migration Procedures				
Configure Resources				
Tool Configuration				
Application Configuration				
Data Configuration				
Develop and Test Data Migration Procedures				
Develop and Test Data Quality Validation Procedures				
Develop and Test Data Cleansing Procedures				
Develop and Test Data Conversion / Transformation Procedures				
Establish Access to Staging and Target Area				
Identify Critical Success Factors for Implementation Phase				
Stage Data				

Data Migration Review Checklist	Status			
	P	C	A	N/A
Create Staging Area				
Populate Staging Area				
Integrate Staged Data				
Validate Staged Data				
Cleanse Data				
Cleanse Data				
Validate Cleansed Data				
Convert / Transform Data				
Convert / Transform Data				
Validate Converted / Transformed Data				
Migrate Data				
Perform Trial Migration				
Validate Results of Trial Migration				
Obtain Approval for Full Migration				
Perform Full Migration				
Validate Full Migration				
Post-Migration				
Operate legacy and target environment in parallel				
Validate parallel operation and data				
Release Data Environments				
Data Migration Closeout				
Document Data Migration Results				
Identify Critical Success Factors for Data Migration Closeout Phase				
Document Data Migration Lessons Learned				
Perform Knowledge Transfer				
Communicate Data Migration and Lessons Learned				

Legend: **P** – Planned

C – Completed

A – Accepted

N/A- Not applicable

Appendix D: Best Practices: Data Migration

Introduction

Most published data migration best practices come from migration-software vendors, indicating that the recommended procedures must be analyzed within the context in which they are written. From the resources, it is clear that no single published best practice should be assumed to be complete.

The best way for any organization to minimize the effects of data migration, including downtime, data loss, and cost, is to employ a consistent, reliable, and repeatable methodology for migrations that incorporates planning, technology implementation, and validation.[49] Such a methodology is often based on a combination of facts, experience, and knowledge gathered from other organizations. Effort is then required to test the policies and procedures that make up the methodology. The use of practices that have been tested, implemented, and published by other organizations as a best practice for achieving a successful data migration can minimize such effort.

A multitude of vendors provide software to support such efforts, and many of them have published best practices to explain how best to accomplish the migration with their software. Such publications serve as the foundation of the practices described herein. These best practices are a starting point for planning a successful data migration effort. Observation of best practices reduces actual planning time as well as testing time.

To be deemed a best practice, a policy or procedure must fall within the following categories:

- It must be published
- It must be a methodology or approach for achieving an objective (not a software tool or specific procedure)
- It must be a successfully implemented methodology, or a methodology explicitly supported by the application of choice

It must be identifiable, independent of any application

The EDM Team at Federal Student Aid recognizes that best practices in data migration provide a benchmark for evaluating individual data migration efforts. The following compendium of best practices was used as background material to develop this best-practice roadmap for Federal Student Aid.

Compendium of Best Practices

Research on Data Migration best practices included the review of

a. Five (5) white papers based on a worldwide data migration survey of 700 IT personnel: *"2006 Best Practices for Data Migration,' 'Simplifying Technology Refresh with Data Migration Software,' and 'The Hidden Costs of Data Migration'* ,

[49] *2006 Best Practices for Data Migration,* page 5

"*The Complete Data Migration Methodology*" (Dulcian, Inc.), and "*Strategic Approach to Data Migration*" by Conversion Services International, Inc.

b. Four (4) articles: "*How to Plan for Data Migration*" (ComputerWorld), "*Best Practices*" (Storage Magazine), "*Taking the Pain Out of Data Migration: Methodology that Works*" (DM Direct), and "*Data Migration: Plan to Succeed*" (DM Direct).

2006 Best Practices for Data Migration

This document "2006 Best Practices for Data Migration" proposes the three-phase methodology shown in Figure 3 to plan, perform, and validate the migration.[50] Softek's methodology provides a useful checklist of which steps should be done in each phase. The first of the Softek documents proposes the methodology shown in Figure 7. The methodology is extensive compared to many other published practices, but is not all-inclusive.

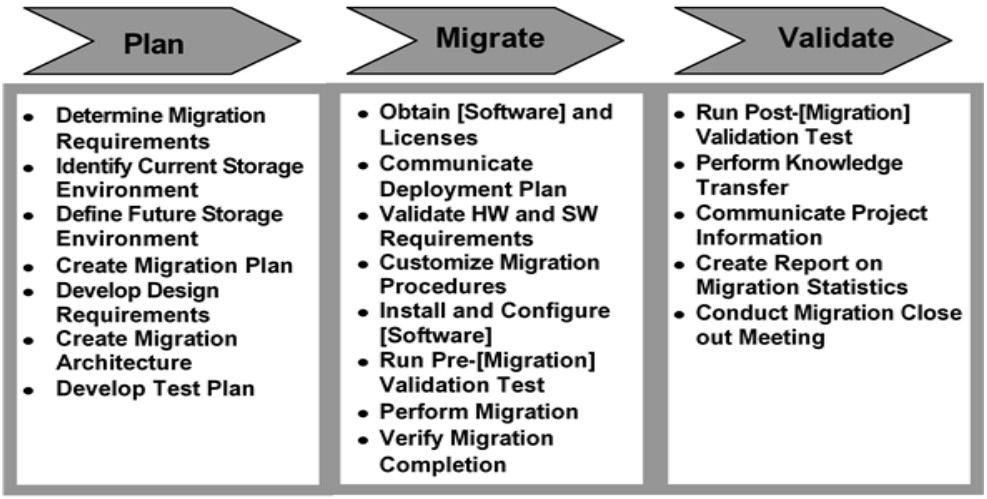


Figure 7: Softek Migration Methodology.

In Phase 1 (**planning**), the team identifies key variables that determine the scope and difficulty of the migration, such as data complexity, current environment, and target environment. The team should consult the business owners of applications and data to help set priorities and tolerances, as well as to elicit any special desires, expectations, or requirements. Based on its findings, the team creates the migration plan and architecture, as well as a validation test plan.

In Phase 2 (**migration**), the team prepares for the migration through a pre-migration validation test on the data, obtains and installs software tools, and migrates the data to the new environment.

In Phase 3 (**validation**), the team validates the migrated data, compiles migration statistics, and documents any lessons learned from the migration effort.

[50] *2006 Best Practices for Data Migration.*

The key milestones within the Softek planning methodology are:

- Determine the migration hardware and software
- Include these design requirements: *Migration Architecture, Hardware and Software Requirements (including acquiring any necessary licenses)*, *Migration Procedures,* and *Deployment and Test Plans*
- Be cognizant that the importance of migration planning increases with the importance of the data to the business and/or the complexity of the data storage (or operational) environment
- Ensure that business owners are included in the planning to help understand the importance of the data and to coordinate the actual migration around other business activities
- Understand the migration requirements are essential when determining schedules, downtime, online vs. offline migration, et cetera

As shown in Figure 7, the Softek Best Practices document describes recommended procedures for migrating and validating the data, including a sample migration plan and a sample design requirement.

Simplifying Technology Refresh with Data Migration Software

This Softek document "Simplifying Technology Refresh with Data Migration Software" asserts that "technology refresh" – the acquisition of new hardware and software technology – is the most common reason for data migration.[51] The document argues that the team should use methodologies and software tools that permit fast, non-disruptive data migration while applications remain online for users.

To enable such online data migration, the document identifies five criteria by which a team should assess methodologies and software tools:

- Performance: How fast the migration occurs and how adequately applications perform during the migration
- Rollback capability: Ability to terminate the migration and restore the original state of the data and systems
- Increase in volume size: Ability of the software and/or methodology to migrate data from a smaller to a larger storage volume
- Heterogeneous source and target hardware: Ability of the software and/or methodology to migrate data between incompatible hardware platforms
- Application downtime: Whether applications using the migrated data can remain online during the migration

[51] *Simplifying Technology Refresh with Data Migration Software*, page 6

The document closes with examples of how Softek's products might facilitate data migration.

The Hidden Costs of Data Migration

Softek's "The Hidden Costs of Data Migration" describes risks to consider when planning not only the migration of the data, but also the target operational system: However, data migration has become increasingly common and increasingly important in both government and private-sector organizations. As a result, IT departments usually schedule data migration on weekends, requiring costly overtime pay for IT staff.

The ideal solution is to perform online data migration during the week while applications are running. The document says that Softek's data migration products are scalable and hardware-independent. Therefore, they permit organizations to migrate data smoothly, accurately, and at a lower cost.

The Complete Data Migration Methodology

The Dulcian white paper provides specific lists, steps, and checkpoints for migration of data to an Online Transaction Processing (OLTP) or Online Analytical Processing (OLAP) system, but its ideas can apply to almost any data migration project. The paper recommends that each step of the methodology should correspond to a specific milestone in the migration process. The main difference between the methodology in this paper and that discussed in other papers and articles is that Dulcian subdivides several steps into two parts:

- The Strategy step is divided into Pre-Strategy and Strategy
- The Analysis step is divided into Pre-Analysis and Analysis
- The Design step is divided into Pre-Design and Design
- The Test step is divided into Pre-Test and Test

Aside from the subdivision of steps and its focus on OLTP/OLAP target systems, Dulcian's methodology corresponds closely to those recommended by the other papers. Dulcian also suggests criteria for choosing a data transformation tool. Such a tool should

- Support flexible reporting
- Generate migration scripts directly from mapping rules
- Include script-scheduling utilities
- Automatically detect integrity violations
- Be cheaper than operation without the tool

Strategic Approach to Data Migration

The central point of the paper by Conversion Services International, Inc, is that data quality analysis and data cleansing should precede the migration itself because they are easier, cheaper, and less error-prone at the source than at the target system. Also recommended prior to the migration is a "proof of concept" trial migration using a small sample of the source data. As a best practice, the paper recommends using automated data profiling tools that provide both automated error detection and interactive examination of data.

Data Migration: Plan to Succeed

This article appears in the January 19, 2007 issue of the *DM Direct* online newsletter, which is published by *DM Review* magazine. It summarizes some of the key ideas and success strategies surrounding data migration. The information is grounded in IT best practices.

Key ideas discussed include online migrations (while keeping applications running and available), offline migrations (requiring application shutdown during the process), and hybrid migrations (combining features of online and offline migrations). Also discussed are host-based migrations, which rely on system software features to migrate data; application-based migrations, which use application features or specialized hardware devices; and array-based migrations, which are used to migrate data from one disk array to another, or from older to newer disks within the same array.

Success strategies discussed include:

- Plan the data migration carefully. Put an experienced project manager in charge, follow best practices, and include testing and back-out procedures.
- Get expert help. If you don't have a data migration expert in-house, bring in an outside data migration specialist to help with the project.
- Define your objectives clearly. Identify your reasons for the migration (technical, financial, et cetera). Determine how much money you can spend on the migration, how much application downtime you can tolerate, and what quantitative benefits you expect.
- Explore different options. Don't simply select the first technology or methodology that comes through the door. Assess how different approaches would achieve your objectives and fit within your cost and resource-availability constraints.
- Get an executive sponsor. A sponsor can confer credibility on your project and inspire people to cooperate.

How to Plan for Data Migration

This article, from the May 21, 2004 issue of *ComputerWorld Magazine,* discusses challenges to data migration efforts. The first part of the article enumerates four common obstacles to data migration:

- Lack of clear definition of data requirements, making it difficult to distinguish conditions between important and unimportant data
- Distributed islands of data caused by individual departments' desire to keep control over their own data and its supporting infrastructure
- Funding constraints
- Lack of expertise in heterogeneous storage environments, with staff or vendor representatives lacking knowledge of technologies relevant to the migration

The article enumerates four guidelines for data migration:

- Classify data in the light of IT requirements and the need to protect data. Describe conditions for data access, retention, and security (including encryption).
- Define migration requirements, including objectives and expected benefits.
- Survey the IT environment, including all infrastructure technology such as networks, bandwidth, and file servers. Compare the existing environment to the requirements for the migration.
- Design the appropriate consolidation or replacement platform to upgrade the current environment so that it effectively handles both the migration itself and the post-migration management of the data.

Taking the Pain Out of Data Migration

This article appears in the November 4, 2005 issue of the *DM Direct* online newsletter, which is published by *DM Review* magazine.

Over 80 percent of data migrations fail, go over budget, or take longer than planned. As a result, they imperil any other IT projects that depend on them. The five main challenges for data migration are:

- Lack of staff with data migration expertise.
- Poor understanding of source systems.
- Target systems that change during the migration effort.
- Target systems that require complex data validations.
- The need to re-verify migration validity after the effort is complete.

With a sound methodology, organizations can manage those challenges. Without a sound methodology, the organization risks data migration failures or project overruns.

Most data migration projects follow a simple four-stage methodology: (1) analyze the source data; (2) extract the data from the source system and transform it for use by the target system; (3) validate the data for the target system; and (4) load it into the target system. However, this methodology does not address the five challenges that bedevil data migration projects. In particular, errors at earlier stages of the process propagate into later stages, requiring costly and time-consuming rework.

These problems result from doing data migration as a serial process. A successful methodology for data migration uses an iterative cycle rather than a series of steps. You should plan "up front" to cycle through all four stages as many times as needed to get the job done correctly. This methodology sees migration as a self-correcting process, improving the result on each pass through the cycle.

Each stage in an iterative methodology provides feedback for all the other stages. Therefore, it's best to use a single, unified toolset for all four stages. Such a unified toolset can take components and interfaces developed at earlier stages and re-use them in later stages, as well as validating and cleansing data more efficiently. When the data meets all requirements of the

target system, the iterative cycle is complete. You can then migrate the data with full confidence in the result.

Best practices include:

- Dealing with high-risk areas first,
- Paying attention to data sources that pose the greatest risk,
- Adapting the migration to any unique features of the target system, and
- Following an audit framework that includes audit trails.

Ideas discussed in this article but not in the others mentioned herein include:

- The standard approach to data migration (analyze, extract/transform, validate, and load) fails to adequately handle the problems usually faced by migration efforts
- Better than the standard approach is an iterative method that repeats the standard process until the migration is complete. The team can validate and fine-tune its techniques, assumptions, and technical infrastructure after each iteration
- When an iteration meets all the constraints of the source systems, requirements of the target systems, and needs of the business, the migration is complete

How to Better Connect Storage to the Business

This article appears in the magazine's monthly "Best Practices" department for November 2006. For data migration, a useful insight from the article is that data migration projects, software, and methodologies must retain historical data and keep it readily available as needed. The rest of the article deals with how to ensure that data storage infrastructure adequately supports the organization's business needs, principally via the use of the Service-Oriented Infrastructure methodology.

Appendix E: Data Migration Project Deliverables

Data Migration Deliverables		Status			
Deliverable	**Description**	**P**	**C**	**A**	**N/A**
Planning					
Project Management Plan					
R / C / A / D	Documentation outlining the Risks, Constraints, Assumptions, & Dependencies associated with proposed data migration.				
Risk Mitigation Plan	Documentation describing the plan for preventing or responding to the risks associated with the data migration; including a matrix demonstrating the likelihood & impact of occurrence of specific risks.				
Project WBS	Documentation showing planned project activities and the time and personnel resources to be applied to performing those activities.				
Data Migration Plan	Plan outlining how the data migration from the data source to the target system is planned for. It includes the plan for ensuring that post-migration data content satisfies the requirements of the target data environment. This document supplements the overall Project Plan and consists of multiple deliverables:				
Data Conversion Plan	The plan for converting the form and content of source data into satisfactory target data.				
Data Quality Metrics	Documentation describing the standards of measurement to be applied to the content data before, during, and after migration. This document includes information regarding the acceptable level(s), if any, of lost data content or meaning as a result of source data undergoing a change in content or form.				
Data Quality Remediation Plan	The plan for correcting any identified data quality issues impacting or resulting from the data migration. *(RE: 3.4.4.4.3 Data Migration Plan)*				
Integration / Reconciliation	The plan for integrating and reconciling all the different source data system into one set of satisfactory target data.				
Data Migration Test Plan	The plan for validating the successful movement of source data to the new target data store.				
Data Archival Strategy	The strategy for archiving historical data that is no longer needed.				

Data Migration Deliverables		Status			
Deliverable	**Description**	**P**	**C**	**A**	**N/A**
Change Management Plan	Documentation showing the process by which version control of project documentation and configuration management will be performed such that all results meet the highest reasonable expectations of quality.				
QA Plan	Documentation showing the process by which the project team shall ensure that all activities are performed such that all results meet the highest reasonable expectations of quality.				
Communications Plan	Documentation showing how information shall be communicated among the members of the project team and between members of the project team and external stakeholders.				
Data Security Plan	Documentation showing how the data security will be implemented and adhered to. This document could reference existing Federal Student Aid Data Security policies and procedures.				
Data Migration Requirements	Documentation describing conditions that must be met in order to deem the data migration successful.				
Parallel Operation Plan	The plan for validating the successful movement of source data to the new target data store through parallel operation of the old and the new system over a defined period of time.				
Analysis & Design					
Data Profile Report	Documentation showing the data quality challenges assessed through the *Data Profiling Report*.				
Data Quality Report (Pre-migration)	Documentation showing how the data quality is assessed and the planned remediation.				
Security Controls	Documentation showing how the data security will be implemented and adhered to.				
Migration Data Architecture	Documentation showing how the data architecture is designed at the logical and physical level through ERDs.				
Staging Area LDM	**Staging Area Logical Data Model:** Graphical representation of the information and business rules of the staging area. (**Note**: *representation should comply with FSA standard modeling methodology*.)				
Staging Area PDM	**Staging Physical Data Model:** Graphical representation of the internal data structures and constraints of the staging area. (**Note**: *representation should comply*				

Data Migration Deliverables		Status			
Deliverable	**Description**	**P**	**C**	**A**	**N/A**
	with FSA standard modeling methodology).				
Staging Area DD	**Staging Data Dictionary**				
Target LDM	**Target Logical Data Model**: Graphical representation of the information and business rules of the target data store. (**Note**: *representation should comply with FSA standard modeling methodology*)				
Target PDM	**Target Physical Data Model**: Graphical representation of the internal data structures and constraints of the target data store. (**Note**: *representation should comply with FSA standard modeling methodology*)				
Target DD	**Target Data Dictionary**				
Data Migration Activity Model	Graphical representation of the activities / functions to be performed during the data migration. (**Note**: *representation should comply with FSA standard modeling methodology*)				
Data Migration CRUD Matrix	Documentation outlining who has access in what capacity to which environment.				
Implementation					
Fully developed and tested Data Migration Procedures	Documentation and code of the data migration procedures and related test results.				
Fully developed and tested Data Validation Procedures	Documentation and code of the data validation procedures and related test results.				
Fully developed and tested Data Cleansing Procedures	Documentation and code of the data cleansing procedures and related test results.				
Fully developed and tested Data Conversion/Transformation Procedures	Documentation and code of the data conversion/transformation procedures and related test results.				
Data Cleansing Report	Documentation of data cleansing findings.				
Data Conversion Report	Documentation of the executed data conversion.				
Trial Migration Results	Documentation of the executed trial data migration(s).				
Acceptance / Approval Documentation	Documentation demonstrating the stakeholder approval of the data migration procedures readiness for deployment to production.				
Full Migration Results	Documentation of the results of the full data migration.				
Parallel Operations Report	Documentation outlining the results of the parallel operation of the old and new data store. This report enables stakeholders to decide whether the parallel operation can be ended.				

Data Migration Deliverables		Status			
Deliverable	**Description**	**P**	**C**	**A**	**N/A**
Close-out					
Data Migration Results	Documentation describing ♦ Statistics of the data migration such as actual data quality and volume measurements, downtime, data loss, etc.; ♦ Unresolved issues; ♦ *[Others]* The actual content of this artifact shall be determined by the overall scope of the data migration. Complex migration efforts would naturally require more extensive reporting, while more basic migrations may require less content.				
Data Migration Lessons Learned	Documentation describing lessons learned during the project that may explain results and contribute to enhancements to future data migration efforts.				
Training Plan	Documentation on the knowledge transfer between the data migration team and the stakeholders take place, including the relevant training material.				
Complete set of all project deliverables	The compiled inventory of all artifacts resulting from the data migration project.				

Appendix F: References

The following sources contributed to the content and/or formatting included herein:

- Front page design and single-line logo drawn from the *Department of Education Federal Student Aid Style Guide.*
- Outline format drawn from the format of the *U.S. Department of Education Enterprise Data Architecture – Enterprise Data Standards and Roadmaps.*
- Creative Computing Solutions, Inc. Enterprise Data Management (Operations) Technical Proposal (August 29, 2006).
- Enterprise Data Management (Operations) Statement of Work (August 25, 2006).
- Burry, Christopher & Mancusi, David. "How to plan for data migration." ComputerWorld (May 21, 2004)
- Softek, Inc. *2006 Best Practices for Data Migration* (White Paper). KnowledgeStorm.com: Softek, Inc., 2006.
- Softek, Inc. *Simplifying Technology Refresh with Data Migration Software* (White Paper). KnowledgeStorm.com: Softek, Inc., January 2006.
- Softek, Inc. *The Hidden Costs of Data Migration* (White Paper). KnowledgeStorm.com: Softek, Inc., April 1, 2006.
- Lewis, William. *Resource Conversion* (MS PowerPoint Presentation). Boston, MA: EMELD, January 9, 2004.
- Peipert, Glenn & Cohen, Lori. *Strategic Approach to Data Migration* (White Paper). WWW: Conversion Services, International, Inc., July 2005.
- Manek, Parul. *Microsoft CRM Data Migration Framework* (White Paper). WWW: Microsoft, Inc., April 2003.
- Damoulakis, James. "Best Practices." *Storage Magazine* (November 2006).
- http://en.wikipedia.org/wiki/Best_Practice
- http://en.wikipedia.org/wiki/Data_conversion
- http://en.wikipedia.org/wiki/Data_Migration
- Purba, Sanjiv. *Handbook of Data Management 1999*. Washington, D.C.: Auerbach, 1999.
- Inmon, W.H. *Data Architecture: The Information Paradigm*. Boston: QED Technical Publishing Group, 1992.
- Parthasarathi, Arvind. "*Taking the Pain Out of Data Migration: Methodology that Works". DMDirect Newsletter* (November 4, 2005).
- Perot Systems, *Integrated Partner Management (IPM) – Data Management Plan* (Spring 2007)